Structural studies of short nucleosome arrays containing linker histone H1

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

im Promotionsprogramm
Biomolecules – Structure, Function and Dynamics
der Georg-August University School of Science (GAUSS)

vorgelegt von
Marco Dombrowski
aus Bad Neuenahr-Ahrweiler, Deutschland

Göttingen, October 2021

Members of Thesis Committee and Examination Board

Members of the thesis advisory committee

Prof. Dr. Patrick Cramer
Department of Molecular Biology
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Holger Stark
Department of Structural Dynamics
Max Planck Institute for Biophysical Chemistry, Göttingen

Dr. Wiebke Möbius
Department of Neurogenetics
Max Planck Institute of Experimental Medicine, Göttingen

Members of the examination board

Prof. Dr. Patrick Cramer (1st referee)
Department of Molecular Biology
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Henning Urlaub (2nd referee)
Bioanalytical Mass Spectrometry Group
Max Planck Institute for Biophysical Chemistry, Göttingen

Further members of the examination board

Prof. Dr. Holger Stark
Department of Structural Dynamics
Max Planck Institute for Biophysical Chemistry, Göttingen

Dr. Wiebke Möbius
Department of Neurogenetics
Max Planck Institute of Experimental Medicine, Göttingen

Prof. Dr. Rubén Fernández-Busnadiego
Institute for Neuropathology
University Medical Center Göttingen

Dr. Johannes Söding
Quantitative and Computational Biology Group
Max Planck Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 01 December 2021

Acknowledgements

I would like to thank Patrick Cramer for giving me the opportunity to carry out my PhD work in his lab. I appreciate the freedom and the support that he grants his coworkers and that foster a thriving and stimulating environment.

I am grateful to Holger Stark and Wiebke Möbius who have provided me with valuable support and discussions in their function as members of my thesis advisory committee.

Thanks goes to Svetlana Dodonova whose help has been invaluable to get me started in the lab and who contributed to the manuscript. Thanks also to Maik Engeholm for helping with sample preparation during the coronavirus lockdown in early 2020 and to Christian Dienemann for help with data acquisition during that time and for continued cryo-EM support.

I would like to thank Henning Urlaub and his team for a flawlessly functioning mass spectrometry facility that helped optimize the workflow for protein purification.

Finally on a more personal level, I want to thank all present and past members of the Cramer lab that have contributed to a productive and supportive atmosphere.

Publications

Parts of this work have been submitted for publication.

M. Dombrowski, M. Engeholm, C. Dienemann, S. Dodonova and P. Cramer (2021), Linker DNA trajectories govern human histone H1 binding to nucleosome arrays.

Author contributions:
M.D. designed and conducted all experiments and data analysis unless stated otherwise. M.E. prepared cryo-EM samples of the 4x177 array. C.D. carried out cryo-EM data acquisition for the 4x177 array, maintained EM facility and advised on microscope setup. S.D. conducted image processing of the 4x177 array data. S.D. and P.C. initially outlined the project. P.C. supervised research. M.D., S.D. and P.C. wrote the manuscript, with input from M.E. and C.D..

The following sections or display items of this thesis are excerpted from Dombrowski *et al.* and marked in the Table of Contents, List of Figures, List of Tables and in the item captions with an asterisk (*):

Section 3.5 Single particle cryo-EM

Text was adapted

3.5.2 Cryo-EM data collection

3.5.3 Cryo-EM data processing

3.5.4 Structural model building

3.5.5 Analysis of linker DNA trajectories†

†Figure 3.1 added

Section 4.1 Reconstitution of tetranucleosome arrays

Figures were adapted

Figures 4.4 top panel correspond to Figure 1a

Figures 4.6 modified after Figure 1b

Section 4.2 Single particle cryo-EM of tetranucleosome arrays

Text, Figures and Tables were adapted

4.2.1 Structural analysis of tetranucleosome arrays containing H1

Figures 4.7 correspond to Figure 1c

Figures 4.8-4.15 correspond to Supplementary Figures 1-8

Tables 4.1-4.4 correspond to Supplementary Tables 1-4

4.2.2 Overall structure of tetranucleosome arrays containing H1

Figure 4.16 corresponds to Figure 2

4.2.3 Nucleosome stacking in solution

Figure 4.17 corresponds to Figure 3

4.2.4 H1 orientation and DNA interactions

Figure 4.18 corresponds to Figure 4

Figure 4.19 corresponds to Figure 5

4.2.5 H1 binding depends on nucleosome repeat length

4.2.6 H1 binding depends on linker DNA trajectory

Figures 4.20-4.22 correspond to Figures 6-8

Text has been expanded and text style, numbering of figures, tables, literature references and cross references between figures and tables have been adapted and may deviate.

Contents

List of Figures

List of Tables

List of Abbreviations

Å	Ångstrom
Amp	ampicillin
β-ME	β-mercapto ethanol
bp	base pair
BSA	bovine serum albumin
cryo-EM	cryo-electron microscopy
cryo-ET	cryo-electron tomography
CTCF	CCCTC-binding factor
CTF	contrast transfer function
Cu	copper
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
g	gram
H1	linker histone H1
H2A	core histone H2A
H2B	core histone H2B
H3	core histone H3
H4	core histone H4
HAT	histone acetyltransferase
HDAC	histone deacetylase
HDM	histone demethylase
HP1	heterochromatin protein-1
IPTG	isopropyl-β-*D*-1-thiogalactopyranoside
Kan	kanamycin
KMT	histone lysine N-methyltransferase
L	liter
M	mol/L
m	milli (10^{-3})
MNase	micrococcal nuclease
μ	micro (10^{-6})
MWCO	molecular weight cutoff
NCP	nucleosome core particle

NDR	nucleosome depleted region
NFR	nucleosome free region
NRL	nucleosome repeat length
PAGE	polyacrylamide gel electrophoresis
PIC	pre-initiation complex
Pol II	eukaryotic RNA polymerase II
PTM	post-translational modification
RNA	ribonucleic acid
rpm	rotations per minute
SDS	sodium dodecyl sulfate
TSS	transcription start site
v/v	volume per volume
Widom-601	Widom and colleagues' clone 601-based nucleosome positioning sequence
w/v	weight per volume

Summary

The DNA inside the eukaryotic nucleus is bound by octamers of histone proteins to form the nucleosome as the basic repeating unit of chromatin. Along the genome, nucleosomes locally form regularly spaced arrays. The relative position of nucleosomes along the DNA can be described by the nucleosome repeat length (NRL) as the sum of nucleosomal DNA and linker DNA connecting neighboring nucleosomes. Whereas short NRL are prevalent near active promoter and enhancer regions, long NRL can be found in transcriptionally repressed heterochromatin. Linker histone H1 is one of the most abundant nucleosome binding proteins and serves to repress transcription. *In vivo*, long NRL are associated with higher H1 content while regions with short NRL are H1 depleted. While well documented in the literature, the underlying mechanisms for the preferential H1 occupancy in long NRL arrays remain elusive.

Here, we present the cryo-electron microscopy structures of H1-bound tetranucleosome arrays with four physiologically relevant NRL. These NRL are characteristic for active promoter and enhancer regions, gene bodies with active transcription and transcriptionally repressed heterochromatin. The structures reveal an overall similar architecture of local nucleosome organization with zig-zagging linker DNA connecting neighboring nucleosomes. Nucleosomes 1 and 3 form a stack while nucleosome 2 and 4 do not stack. Nucleosomes that do not stack have H1 bound whereas H1 binding to the stacked nucleosomes changes with NRL. Exit and entry DNA trajectories in stacked nucleosomes deviate in short NRL arrays and likely preclude H1 binding, but DNA trajectories successively relax with increasing NRL and approach optimal conditions for both stacked nucleosomes in long NRL arrays.

The structures presented here have important implications for understanding transcription in the context of nucleosome arrays. Our findings indicate that H1 binding may be destabilized in short NRL and suggest an alternative mechanism that contributes to H1 depletion or maintenance of H1 depletion. Further, the stable binding of H1 to long NRL arrays can explain the repressive nature of heterochromatin. The data presented here provide a basis for understanding how higher-order chromatin structure influences binding of chromatin factors to shape chromatin function.

Introduction

1.1 The eukaryotic genome is organized as chromatin

In the late nineteenth century, a phosphorous rich acidic substance, then named nuclein (Workman and Abmayr, 2014; Dahm and Miescher, 2005), and a protein component called histone (Workman and Abmayr, 2014; Kossel, 1884), were discovered in nuclei. Around the same time, this nucleoprotein structure was found to be stainable by basophilic dyes and was thus named chromatin (Workman and Abmayr, 2014). The important function of nucleic acids became clearer only later in the middle of the twentieth century with the discovery of deoxyribonucleic acid (DNA) as the component carrying genetic information (Avery et al., 1944) and the determination of the double helical structure of DNA (Franklin and Gosling, 1953; Watson and Crick, 1953) which provided insights into how genetic information could be copied. Fractionation of the basic histone proteins contained in nuclei revealed two major classes of histones (Stedman and Stedman, 1951), now called core histones H2A, H2B, H3 and H4, and linker histones H1. Further work revealing that H2A and H2B can form dimers (Kelley, 1973) and that H3 and H4 can form tetramers (Kornberg and Thomas, 1974) together with electron microscopy studies showing chromatin fibers as "beads on a string" (Olins and Olins, 1974) gave rise to the model of DNA wrapped around an octamer of histone proteins comprised of an H3-H4 tetramer with two H2A-H2B dimers bound peripherally by one linker histone as the repeating unit of chromatin (Kornberg, 1974). Traditionally, the term 'nucleosome core particle' (NCP) is used to describe the histone octamer and bound DNA whereas 'nucleosome' refers to NCP with linker DNA (Luger et al., 1997). However, for the sake of readability only the term 'nucleosome' will be used in this thesis.

1.2 Nucleosomes are the basic unit of chromatin

DNA wraps around an octamer of histone proteins to form the nucleosome as the basic unit of chromatin. Along the genome, nucleosomes are connected by linker DNA. Nucleosomes occupy most of the genomic DNA (Oberbeckmann et al., 2019). Besides packaging the genome into the constraints of the nucleus, nucleosomes act as a signaling platform and as

important regulators of nuclear processes like DNA replication, transcription and repair (Workman and Abmayr, 2014; Lai and Pugh, 2017). Chromatin represses the initiation activity of the general transcription machinery (Laybourn and Kadonaga, 1991) and also slows down transcription elongation. By modifying nucleosome positions and dynamics, gene activity can thus be regulated in an effective way (Li et al., 2007).

1.2.1 The structure of the nucleosome

The core histones H2A, H2B, H3 and H4 consist of circa 100–130 amino acids. Three α-helices $\alpha1$ to $\alpha3$ linked by linkers L1 and L2 make up the histone fold (Fig. 1.1a) through which H2A-H2B and H3-H4 form antiparallel dimers (Arents et al., 1991) (Fig. 1.1b). The high resolution crystal structure of the nucleosome containing recombinantly expressed *Xenopus laevis* core histones provided key insights into nucleosome function (Luger et al., 1997) (Fig. 1.1c). In the nucleosome, two H3-H4 dimers form the H3-H4 tetramer through a four-helix bundle involving the H3 histone folds and the two H2A-H2B dimers interact with the H3-H4 tetramer through a four-helix bundle between the H2B and H4 histone folds (Luger et al., 1997).

Roughly 1.7 left handed turns of DNA corresponding to 145–147 base pairs (bp) of DNA wrap around the octamer (Luger et al., 1997). The major groove of the central base pair at the pseudo two-fold symmetric axis (dyad axis) of the nucleosome is defined as superhelix location 0 (SHL 0) and serves as a reference for the rotational orientation of the DNA (Luger et al., 1997). Thus, positions on the nucleosome can be described using the range of SHL -7 to SHL +7 with integer numbers indicating the major groove facing the histone octamer and half numbers like SHL+0.5 indicating the minor groove facing the histone octamer (Fig. 1.1c). Direct contacts between histones and DNA mostly involve the DNA phosphodiester backbone and generally occur where the minor groove is facing the octamer (Luger et al., 1997). Major interaction sites are the $\alpha1\alpha1$ helices and the two L1L2 loops of the antiparallel histone fold dimers (Luger et al., 1997) (Fig. 1.1b). Additional interactions with DNA occur through histone fold extensions such as the H3 αN helix that contact exit and entry DNA (Luger et al., 1997).

The flexible N terminal tails of H2B and H3 exit the nucleosome core between the two DNA superhelices while the N terminal tails of H2A and H4 run over the DNA minor groove (Luger et al., 1997). From there, the tails extend flexibly into the solvent (Luger et al., 1997; Davey et al., 2002) and are implicated in interactions with linker DNA and with other nucleosomes in the context of chromatin (Dorigo et al., 2003) and serve as

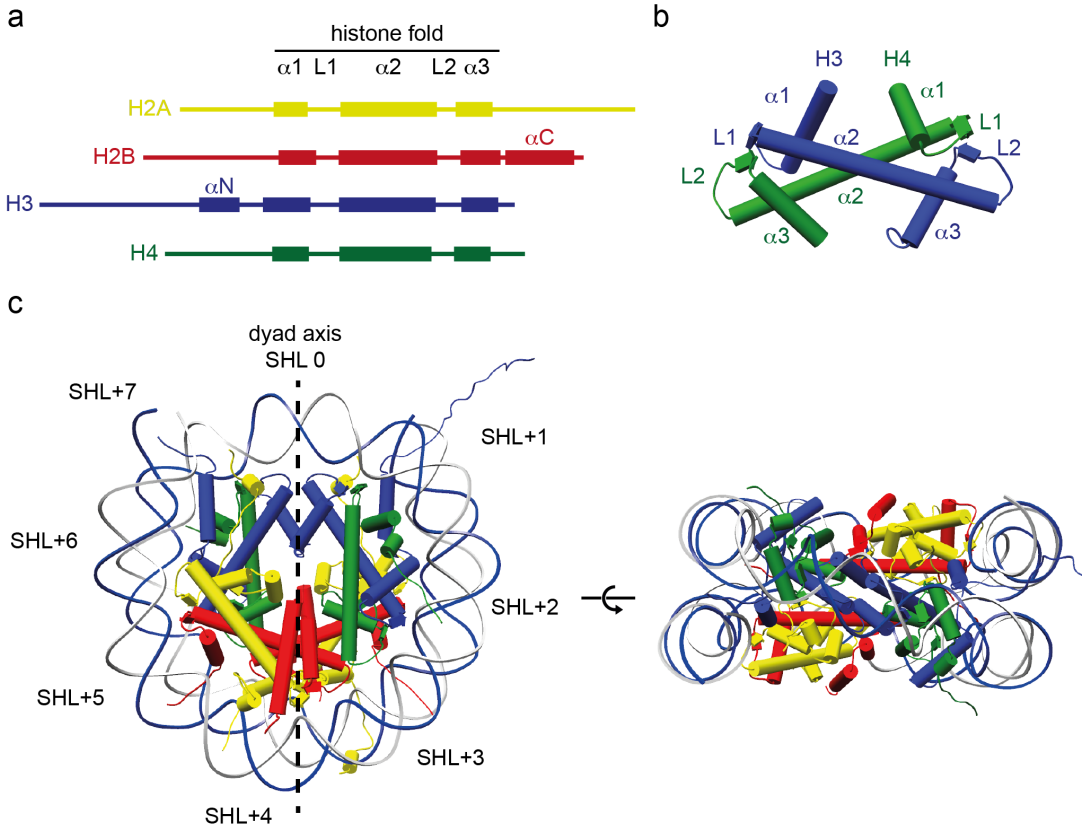

Figure 1.1: The structure of the nucleosome. **(a)** The four core histones H2A (yellow), H2B (red), H3 (blue) and H4 (green) are composed of the histone fold that consists of 3 α-helices $\alpha1$ to $\alpha3$ linked by linkers L1 and L2. **(a)** Dimerization of histones H2A-H2B and H3-H4 through the histone fold domains is depicted here with the H3-H4 dimer. Through this arrangement, the $\alpha1\alpha1$ and the two L1L2 interaction surfaces for DNA contact are formed (chains A and B from PDB ID 1AOI). **(c)** In the crystal structure of the nucleosome containing recombinantly expressed *Xenopus laevis* core histones and modified human α-satellite DNA, 1.7 turns of DNA containing 146 base pairs wrap around the core histone octamer. The nucleosome has pseudo two-fold symmetry. The major groove of the base pair centered on the dyad axis faces the octamer and is defined as superhelix location (SHL) 0. From there in both directions, positions along the nucleosomal DNA are described by SHL-7 to SHL+7 where the major groove faces the octamer. PDB ID 1AOI (Luger et al., 1997). DNA in white and blue, H2A in yellow, H2B in red, H3 in blue, H4 in green. Panels **a** and **b** after Workman and Abmayr (2014).

signaling hubs for the regulation of nuclear processes through various post-translational modifications (Jenuwein and Allis, 2001; Allis and Jenuwein, 2016)

The core histones are highly conserved across species (Baxevanis and Landsman, 1998) and the structures of *Xenopus laevis* (Luger et al., 1997; Davey et al., 2002), chicken (Harp et al., 2000), *Saccharomyces cerevisiae* (White et al., 2001) and human (Tsunaka et al., 2005) nucleosomes are very similar. Among the conserved structural features of the nucleosome is the acidic patch formed by glutamate and aspartate residues of histones H2A and H2B on the surface of the octamer (Luger et al., 1997). The acidic patch has been implicated in interactions between adjacent nucleosomes (Luger et al., 1997; Kalashnikova et al., 2013a). Additionally, other chromatin factors like chromatin remodelers and histone modifiers typically interact with the nucleosome acidic patch using variations of the arginine anchor motif (McGinty and Tan, 2016, 2021).

Several core histone variants exist that alter nucleosome structure and have implications for chromatin structure and transcription regulation (McGinty and Tan, 2015). The H2A variant H2A.Z changes octamer surface to include a metal ion, potentially affecting interactions between nucleosomes and recruitment of chromatin factors (Suto et al., 2000). Additionally, H2A.Z destabilizes the H2A.Z-H2B dimer interface to H3-H4 (Suto et al., 2000). The H3 variant CENP-A that is prevalent at centromeric nucleosomes contains shorter αN helix that binds less DNA leading to more open conformation of entry and exit DNA (Dechassa et al., 2011; Tachiwana et al., 2011).

1.2.2 Nucleosome assembly

Histone genes

The replication dependent core histone genes are encoded in the large *HIST1* cluster on chromosome 6 and the smaller *HIST2* and *HIST3* clusters on chromosome 1 in humans (Albig and Doenecke, 1997) and are transcribed during S-phase. Altogether, these clusters encode for more than 10 copies of each of the four core histones (Marzluff et al., 2002). While all H4 genes encode the same protein and all H3 genes encode 3 different proteins, there is more variation within the H2A and H2B genes (Marzluff et al., 2002). Histone genes typically lack introns (Pandey et al., 1990) and the mature messenger RNA (mRNA) ends with a stem-loop rather than a poly(A) tail (Dominski and Marzluff, 1999). The processing and cleavage of the 3' end is carried out by the stem-loop binding protein SLBP and the U7 snRNP which binds the purine rich histone downstream element (HDE)

(Dominski and Marzluff, 1999). Other replication independent replacement histone variants such as H3.3 and H2A.Z are encoded by spliced and polyadenylated mRNAs (Albig et al., 1995; Hatch and Bonner, 1990).

Nucleosome assembly

Together with *in vitro* experiments showing that H2A-H2B exists as dimers (Kelley, 1973) and that H3-H4 dimers are in equilibrium with H3-H4 tetramers (Baxevanis et al., 1991), the structure of the nucleosome led to a model of nucleosome assembly in which first an H3-H4 tetramer or two H3-H4 dimers are deposited onto DNA, leading already to the coordination of more than half of the full nucleosomal DNA. Then, two H2A-H2B dimers are added to complete the octamer and bind the remaining nucleosomal DNA. Disassembly is thought to happen in reverse. During assembly and disassembly, histone chaperones are proposed to bind to free histones and prevent aggregation and unspecific interactions with DNA (Laskey et al., 1978; Aguilar-Gurrieri et al., 2016; Hammond et al., 2017).

1.2.3 Nucleosomes and transcription

The first step in transcription by RNA polymerase II (Pol II) is the assembly of the pre-initiation complex on promoter DNA in a process called initiation (Hantsche and Cramer, 2016; Schilbach et al., 2017; Aibara et al., 2021). Nucleosomes block transcription initiation (Lorch et al., 1987) and have to be removed for initiation to occur (Lorch and Kornberg, 2017). While transcription through the nucleosome is possible *in vitro*, it is substantially slower than on naked DNA as the nucleosome represents a major barrier for Pol II (Izban and Luse, 1992; Farnung et al., 2018). On the way to the nucleosome dyad, Pol II transiently stalls near major contact sites between DNA and histones whereas the resistance encountered by Pol II decreases drastically after passing the dyad (Kujirai et al., 2018; Chen et al., 2019). During nucleosome transcription, Pol II is unable to actively detach DNA from the histone octamer and is thought to act as a ratchet trapping transiently unwrapped states (Hodges et al., 2009). The concerted and regulated action of elongation factors and Pol II is necessary to facilitate transcription through chromatin (Vos et al., 2020). During nucleosome transcription, histone chaperones and chromatin remodelers are involved in preventing histone loss and maintaining chromatin (Smolle et al., 2012; Skene et al., 2014; Farnung et al., 2021).

1.2.4 Nucleosome positioning *in vivo*

Nucleosomes are thought to regulate transcription by physically blocking access to DNA (Lai and Pugh, 2017). It is therefore interesting to study where nucleosomes are positioned along the genome, how this relates to transcriptional activity and how this changes during transcriptional activation and deactivation. The relative position of nucleosomes along the genome can be described by the nucleosome repeat length (NRL) as the sum of the base pairs of the nucleosome bound DNA and the linker DNA connecting it to the next nucleosome. Several methods have been developed to map nucleosome positions (Lieleg et al., 2015). Strategies include but are not restricted to controlled digestion with micrococcal nuclease (MNase) to generate short genomic fragments containing one or just a few nucleosomes that can be used in tiling arrays (Yuan et al., 2005) or sequenced and mapped back onto the genome (Baldi et al., 2018b).

Genome wide mapping of nucleosomes in yeast revealed a stereotypical pattern of nucleosome positions at promoter regions (Yuan et al., 2005). There, a nucleosome free region (NFR, sometimes referred to as nucleosome depleted region NDR) is flanked by two well positioned nucleosomes, the -1 nucleosome upstream and the +1 nucleosome downstream (Yuan et al., 2005) (Fig. 1.2 (Lai and Pugh, 2017)). A phased nucleosome array extends into the gene body, that is, nucleosomes are spaced regularly and are aligned to a common reference (Yuan et al., 2005). The average NRL in yeast is around 167 bp (Holde, 1998). Interestingly, transcriptionally active regions have shorter NRL than inactive regions (Chereji et al., 2018; Ocampo et al., 2016).

The presence of phased arrays near the transcription start site is conserved also in plants and animals (Lai and Pugh, 2017). But whereas most nucleosomes in *Saccharomyces cerevisiae* are well positioned with respect to the underlying DNA (Yuan et al., 2005), nucleosomes tend to occupy less defined positions in animals (Valouev et al., 2011) making measurements of nucleosome spacing and NRL more difficult (Baldi et al., 2020). In single cell MNase seq studies of human cells, it was found that nucleosome spacing is more uniform in single cells and that this information is lost when looking at bulk data and mononucleosome fragments (Lai et al., 2018). Interestingly, heterochromatic regions and inactive promoter regions exhibit ill defined positioning but highly regular spacing of nucleosomes (Lai et al., 2018). However, nucleosomes near transcription start sites of actively transcribed genes are well defined but less regularly spaced (Lai et al., 2018). Nucleosome arrays within actively transcribed exhibiting H3K36me3 also appeared to be

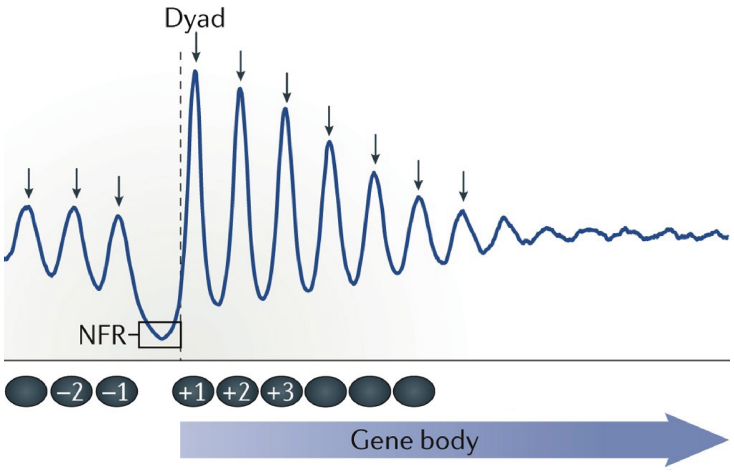

Figure 1.2: Stereotypical arrangement of nucleosome arrays at the 5' end of genes near active promoters in yeast. Most nucleosomes in yeast are well positioned and arranged in a stereotypical way at active gene promoters. A nucleosome free region (NFR, sometimes referred to as nucleosome depleted region NDR) is flanked by the well positioned -1 and +1 nucleosome and followed by a phased array of nucleosomes extending into the gene body. Adapted from Lai and Pugh (2017).

more regularly spaced (Lai et al., 2018). Similar findings were also reported for *Drosophila* where a correlation between short NRL and higher transcriptional activity was also observed (Baldi et al., 2018b) as was previously reported for human cells (Valouev et al., 2011).

While it is thus still elusive how nucleosome spacing influences transcription and vice versa, it is becoming clear that regular arrays of nucleosomes are present throughout the genome (Baldi et al., 2018b, 2020; Lai et al., 2018). Intriguingly, NRL and transcriptional activity also seem to be related. Short NRL are more prevalent in expressed genes whereas long NRL are found in silent regions (Baldi et al., 2018b; Chereji et al., 2018; Ocampo et al., 2016; Valouev et al., 2011; Lai et al., 2018). Taking the average genome wide NRL of a wide range of eukaryotic cells, a preferred linker DNA length quantization of integer multiples of 10 ($10n$, $n \in \mathbb{N}$) was observed (Widom, 1992).

1.2.5 DNA sequence influences nucleosome positioning

Studies have shown that in the absence of other factors, DNA sequence can influence nucleosome positions (Struhl and Segal, 2013). As nucleosomal DNA geometry deviates substantially from free B-DNA (Luger et al., 1997) it is not surprising that due to differences in bendability (Drew and Travers, 1985), different DNA sequences show different affinities for octamer binding (Thåström et al., 1999). Bendable dinucleotides containing dA and dT are preferentially positioned at half SHLs where the DNA minor groove faces and directly interacts with the octamer and dinucleotides containing dG and dC are pref-

erentially positioned at integer SHLs with the major groove away (Satchwell et al., 1986).

Widom and colleagues exploited these intrinsic biophysical properties of DNA sequences and employed iterative *in vitro* selection for nucleosome binding followed by amplification to generate DNA sequences with high affinity for histone octamer binding (Lowary and Widom, 1998). Clone 601 and derivations thereof (Widom-601) are widely used in studies due to their well defined nucleosome positioning ability.

Long runs of dA or dT exclude nucleosomes *in vitro* due to their stiffness (Struhl and Segal, 2013; Nelson et al., 1987). However, the precise mechanism of nucleosome depletion on poly(dA) tracts *in vivo* is controversial (Barnes and Korber, 2021). Recent work revealed that chromatin remodelers may be able to read out biophysical properties of DNA to direct nucleosome positioning (Oberbeckmann et al., 2021a). Further work is necessary to dissect the precise contribution of DNA sequence to nucleosome positioning.

1.2.6 Chromatin remodelers position nucleosomes *in vivo*

The biophysical properties of DNA alone have been shown to be insufficient to explain the nucleosome positions observed *in vivo* (Zhang et al., 2009). Rather, a family of proteins called ATP-dependent chromatin remodeling enzymes (remodelers) regulates nucleosome positioning by catalyzing nucleosome eviction, sliding, spacing or histone variant incorporation (Flaus et al., 2006; Becker and Workman, 2013; Clapier et al., 2017). Remodelers can broadly be categorized into the four subfamilies ISWI, CHD, SWI/SNF and INO80 based on sequence differences in their catalytic ATPase domains (Flaus et al., 2006; Clapier et al., 2017). All remodelers except for the CHD subfamily are multiprotein complexes. Most remodelers can slide nucleosomes while nucleosome eviction is only catalyzed by RSC and Swi/Snf of the SWI/SNF subfamily (Clapier et al., 2017). Nucleosome spacing is carried out by CHD family remodelers, ISW1a and ISW2 of the ISWI subfamily and INO80 of the INO80 subfamily (Clapier et al., 2017).

Remodelers with sliding activity are thought to work together with barrier factors to direct nucleosome positioning (Zhang et al., 2009; Oberbeckmann et al., 2021b; Krietenstein et al., 2016; Wiechens et al., 2016; Baldi et al., 2018a). One such barrier factor in mammals is the architectural protein CTCF (Wiechens et al., 2016) but it is also conceivable that nucleosomes themselves serve as barriers. The spacing of nucleosomes would then be brought about by sliding remodelers that contain an intrinsic structural feature termed the ruler that reads out distance between nucleosomes (Yamada et al., 2011). A recent survey on a variety of sliding remodelers found evidence for such rulers in

remodelers with known spacing activity (Oberbeckmann et al., 2021b). The stereotypical organization near active promoters is currently thought to start with the generation of the NDR by a combination of poly(dA) stretches repelling nucleosomes and eviction activity by RSC through binding the poly(dA) site and possible associated barrier factors, followed by INO80 positioning the +1 nucleosome through DNA sequence readout (Lieleg et al., 2015; Krietenstein et al., 2016). Spacing remodelers such as ISW1a and Chd1 would then generate spaced arrays (Lieleg et al., 2015; Krietenstein et al., 2016). Interestingly, these observations suggest that, in the absence of perturbations through barrier factors or recruitment of nucleosome evicting remodelers, regularly spaced nucleosome arrays would be generated across the genome.

1.2.7 Post-translational modifications of histone tails

Nucleosomes are heavily post-translationally modified in a regulated way mainly in the N terminal tails of the core histones (Jenuwein and Allis, 2001; Allis and Jenuwein, 2016). Early work on histone acetylation and methylation linked histone tail modifications to transcription regulation (Allfrey et al., 1964). This hypothesis was further corroborated by the identification of histone acetyl transferases (HAT) that functions as a transcriptional coactivators (Brownell et al., 1996; Kuo et al., 1996; Mizzen et al., 1996; Yang et al., 1996; Ogryzko et al., 1996) and the discovery of a transcriptional repressor with histone deacetylase (HDAC) activity (Taunton et al., 1996). The observation of the bromodomain binding to acetylated lysines (Dhalluin et al., 1999) triggered the histone code hypothesis suggesting that specific histone tail modifications or combinations thereof recruit specific factors such as chromatin remodelers that can read these modifications and carry out their biological function in a targeted way (Strahl and Allis, 2000; Jenuwein and Allis, 2001).

Chromatin states would in that way be marked by writers with a specific combination of post-translational histone modifications that is read out by readers and undone by erasers. Promoters are marked by H3 lysine 4 trimethylation and possibly H3 lysine 9 acetylation (H3K4me3 and H3K9ac), while enhancers, DNA elements that enhance activity of promoters typically far away on the linear genome, carry an H3 lysine 4 monomethylation and H3 lysine 27 acetylation (H3K4me1 and H3K27ac) signature (Dunham et al., 2012; Rada-Iglesias et al., 2011). Transcribed gene bodies are marked with H3 lysine 36 trimethylation (H3K36me3) (Dunham et al., 2012). Heterochromatin is characterized by high levels of H3 lysine 9 trimethylation (H3K9me3) (Dunham et al., 2012) which in-

hibits histone acetylation by p300 and recruits heterochromatin protein-1 (HP1) (Stewart et al., 2005). Recruitment of HP1 is abrogated by H3 serine 10 phosphorylation (H3S10p) (Fischle et al., 2005), highlighting the possibility of interfering effects from neighboring modifications.

In addition to the *trans* read out by chromatin factors, histone modifications can also act directly in *cis* by modulating internucleosome or intranucleosome interactions (Pepenella et al., 2014). The H3 N terminal tail exits the octamer near the DNA entry and exit sites of the nucleosome (Luger et al., 1997) and interact preferentially with linker DNA (Angelov et al., 2001). There are indications that post-translational modifications change H3 tail dynamics, modifying contacts with the nucleosome, and that this modulates binding of other factors by making recognition sites more accessible (Stützer et al., 2016; Morrison et al., 2018). In the crystal structure of the nucleosome, the H4 tail has been observed to interact with the acidic patch of another nucleosome particle (Luger et al., 1997). This interaction was later shown to influence internucleosome interactions and thus direct chromatin folding (Dorigo et al., 2003). Neutralization of the positive charge by acetylation of lysines abrogates this interactions and modulates chromatin fiber folding (Zhang et al., 2017; Allahverdi et al., 2011).

1.3 Linker histone H1 binds to nucleosomes

Linker histone H1 is one of the most abundant proteins in the eukaryotic nucleus. H1 binds to the nucleosome to form the next recurring structural unit of chromatin that is sometimes referred to as the chromatosome (Simpson, 1978). The eleven H1 variants in mammals are products of gene duplication and are thus paralogs (Izzo et al., 2008). A unified nomenclature was proposed according to which they are called H1.0 to H1.10 (Talbert et al., 2012) and will be adhered to throughout this thesis. There exist seven somatic subtypes (H1.0, H1.1 to H1.5, H1.10), three testis specific subtypes (H1.6, H1.7 and H1.9) and one oocyte specific subtype (H1.8) (Hergeth and Schneider, 2015; Fyodorov et al., 2018). Of the somatic variants, H1.0 is enriched in terminally differentiated cells while the others are ubiqitously expressed (Fyodorov et al., 2018). The genes of somatic variants H1.1 to H1.5 and the testis specific H1.6 are located in the *HIST1* cluster on chromosome 6 (Millán-Ariño et al., 2016).

1.3.1 The structure of H1

Linker histones in mammals consist of about 200 amino acids and have a conserved three part architecture consisting of a short unstructured N terminal domain (NTD), followed by a globular domain composed of 70 amino acids and finally a 100 amino acid long unstructured C terminal domain (CTD) (Fig. 1.3a). The crystal structure of the globular domain of chicken H1.0 (called H5 in birds) revealed a winged helix motif consisting of three α helices α1 to α3 followed by a two-stranded β-sheet (Ramakrishnan et al., 1993). The amino acid sequence of the globular domain of H1.0 to H1.5 including positively charged residues implicated in nucleosome binding are highly conserved (Fig. 1.3a). In contrast to this, linker histone NTD and CTD show more variation between subtypes.

Surprisingly, high resolution insight into H1 binding to nucleosomes was obtained only recently (Zhou et al., 2013, 2015). *Drosophila* encodes for one H1 and an additional variant BigH1 with a larger N terminal tail is expressed during development (Perez-Montero et al., 2013). *Drosophila* H1 was shown by nuclear magnetic resonance (NMR) to bind to the nucleosome off-dyad asymmetrically with the α3 helix near the dyad and contacted 10 bp of one linker DNA (Zhou et al., 2013). However, by a combination of X-ray crystallography and NMR, chicken H1.0 was observed to bind on the nucleosome dyad axis and contact both linker DNAs (Zhou et al., 2015). This difference in binding mode was later demonstrated to be due to the differing position of five positively charged residues between *Drosophila* H1 and chicken H1.0 (Zhou et al., 2016). The human somatic variants H1.5 (Bednar et al., 2017) and H1.0, H1.4 and H1.10 (Zhou et al., 2021a) have been studied and all bind to the nucleosome in a very similar way, termed on-dyad, contacting nucleosomal DNA and the two linker DNAs (Fig. 1.3b), leading to subtly different linker DNA dynamics (Zhou et al., 2021a). Even though the long CTD contributes substantially to the high affinity binding to nucleosomes (Hendzel et al., 2004; White et al., 2016), and was shown to preferentially associate with one linker DNA (White et al., 2016; Bednar et al., 2017; Zhou et al., 2021a), it appears to remain largely disordered upon binding (Caterino and Hayes, 2011; Bednar et al., 2017; Zhou et al., 2021a).

1.3.2 H1 binding to chromatin

Studies using fluorescence recovery after photobleaching (FRAP) found that linker histone H1 is transiently associated with chromatin *in vivo* with continuous exchange of H1 occuring in both euchromatin and heterochromatin (Lever et al., 2000; Misteli et al., 2000).

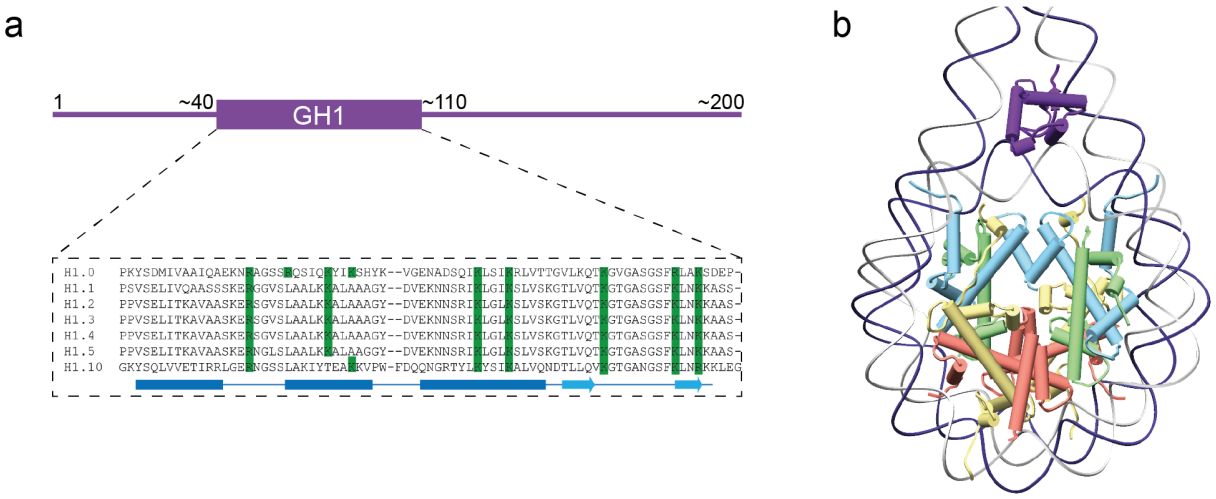

Figure 1.3: The structure of linker histones. (a) Linker histones H1 consist of a 70 amino acid large globular domain (GH1) flanked by a short unstructured N terminal tail and a long unstructured C terminal domain. The sequence of the globular domain is conserved among the somatic H1 variants H1.0, H1.1 to H1.5 and H1.10 including positively charged residues that are important for binding to the nucleosome (highlighted in green). GH1 consists of a winged helix domain characterized by three α helices followed by a two-stranded β sheet (secondary structure element cartoon in blue). Protein sequences were taken from UniProt and aligned using PROMALS3D (Pei et al., 2008). **(b)** The human somatic H1 variants H1.0, H1.4, H1.5 and H1.10 bind to the nucleosome on the dyad axis contacting nucleosomal DNA and the two linker DNAs (Bednar et al., 2017; Zhou et al., 2021a). Even though H1 is bound asymmetrically and lopsided, this mode is referred to as on-dyad. Shown exemplarily here is the cryo-electron microscopy structure of the human somatic H1 variant H1.4 bound to the nucleosome (PDB ID 7K5Y Zhou et al. (2021a)) with the core histones in standard colors and H1 in purple.

The relatively long residence time of several minutes depends on H1 CTD (Lever et al., 2000; Hendzel et al., 2004). The identification of populations with distinct mobilities that appeared to depend on H1 acetylation indicated that PTMs might regulate linker histone binding (Misteli et al., 2000). Indeed, several modifications have been identified that are linked to increased H1 dynamics or its eviction from chromatin (Izzo and Schneider, 2016; Fyodorov et al., 2018).

It has been shown that H1 bound nucleosomes are generally refractory to nucleosome sliding even though specialized complexes may be able to remodel the H1-nucleosome complex (Maier et al., 2008; Zhou et al., 2021a). Several proteins factors have been proposed to induce H1 eviction from chromatin. Direct competition of the pioneer transcription factor FOXA1 with H1 has been proposed as it contains a similar winged helix domain and binds to a similar site on the nucleosome (Cirillo et al., 1998, 2002). Similar modes of action have been proposed for poly(ADP-ribose) polymerase 1 (PARP1) and proteins belonging to the high mobility group (HMG) family (Krishnakumar et al., 2008; Postnikov and Bustin, 2016). Recent work uncovered a direct link between core histone acetylation and H1 eviction and showed that both are necessary to induce transcription from H1 bound promoters (Shimada et al., 2019).

1.3.3 H1 distribution along the genome

How different H1 paralogs are distributed along the genome has remained elusive as the detection of specific subtypes is hindered by lack of specific antibodies and the more diffuse pattern of H1 binding (Millán-Ariño et al., 2016). Nevertheless, studies using chromatin immunoprecipitation (ChIP) combined with sequencing (ChIP-seq) of hemagglutinin-tagged (HA-tagged) recombinantly knocked in H1 variants (Millán-Ariño et al., 2014) or employing the DNA adenine methyltransferase identification (DamID) technique (Izzo et al., 2013) provided some insight into the differential distribution of H1 variants.

Studies in *Drosophila* showed similar levels of H1 throughout the genome in both heterochromatin and euchromatin (Braunschweig et al., 2009). However, there and in the human cancer cell line MCF7 total H1 is depleted in proximity of transcription start sites (TSS) of active genes ("H1 valley") and other regulatory regions (Krishnakumar et al., 2008). In mouse embryonic stem cells, H1.2 and H1.3 are inversely correlated with the active histone mark H3K4me3, enriched at the repressive histone mark H3K9me3 and at major satellite sequences but depleted at active promoters (Cao et al., 2013). In a survey of somatic variants H1.1 to H1.5 in human cells, depletion of H1 from the TSS scaled with gene activity while H1 was broadly distributed across the gene body (Izzo et al., 2013). In a different human cancer cell line, depletion of total H1 extended beyond the NDR and suggested that transcription depends on removal of H1 from promoter regions (Millán-Ariño et al., 2014). Here the authors also observed a correlation between H1.2 and gene repression while it appears to be depleted at the TSS of inactive genes (Millán-Ariño et al., 2014). In another study H1.10 was found to be enriched at coding regions and Pol II associated regions while H1.0 was associated with repetitive DNA sequences in the nucleolus such as ribosomal DNA (Mayor et al., 2015). The variant H1.5 was found to be enriched at splice sites implicating it in splicing regulation (Glaich et al., 2019). However, enrichment of specific H1 variants at specific genomic loci does not appear to be conserved across different cell lines and might be related to the relative variant expression levels (Millán-Ariño et al., 2014) suggesting that insights might not be directly applicable to other cell types or developmental conditions.

While much work remains to be done to improve our understanding of the distribution of specific H1 subtypes, some common themes such as depletion of total H1 near active promoters and regulatory sequences like CpG islands and CTCF sites and the enrichment of certain subtypes at specialized genomic regions have emerged (Millán-Ariño

et al., 2016).

1.3.4 H1 content correlates with nucleosome repeat length

Chromatin reconstituted *in vitro* in the absence of linker histone H1 shows short spacings and titration of H1 leads to an increase in NRL (Blank and Becker, 1995; Eggers and Becker, 2021). Interestingly, different H1 variants appear to induce different spacings (Öberg et al., 2012). Cations with different valencies had a similar effect on increasing NRL with higher concentration that depended on their total charge, suggesting that the propensity for electrostatic neutralizations might influence the NRL (Blank and Becker, 1995). But how this effect translates mechanistically is unclear.

Similar observations were made *in vivo*. Knock out studies targeting three major somatic variants in mouse embryonic stem cells led to reduction in H1 levels by circa 50% that was concomitant with a global reduction of NRL (Fan et al., 2005). Conversely, overexpression of two H1 variants leading to circa 1.2 fold increase in H1 expression caused an increase in NRL (Gunjan et al., 1999). Comparisons across several organisms, tissue types and cell types revealed an almost linear relationship between H1 to nucleosome stoichiometry and NRL (Woodcock et al., 2006). Thus, while the correlation between H1 content and NRL exists and higher H1 content seems to be associated with longer NRL, the underlying molecular mechanisms for this phenomenon remain elusive.

1.3.5 H1 function in transcription

While initially linker histone H1 was thought to be generally repressive for transcription, recent work revealed that H1 carries out diverse functions that modulate gene activity rather than globally inhibit it (Fan et al., 2005; Fyodorov et al., 2018).

H1 regulates epigenetic marks

Part of H1 repressive activity may stem from inhibiting core histone acetylation, marks usually associated with transcriptional activity and open chromatin (Herrera et al., 2000; Sun et al., 2015). H1 has also been shown to inhibit SET7/9, the HMT that catalyzes the core histone methylation of H3K4me3 that is characteristic for active promoters (Yang et al., 2013). How H1 carries out its inhibitory effects on histone modifying enzymes is unknown. Studies indicate that H1 binding to the nucleosome may alter H3 tail dynamics either directly or indirectly through the CTD and thereby modulate its accessibility

(Stützer et al., 2016; Morrison et al., 2018; Hao et al., 2020; Zhou et al., 2021a).

Additionally, recruitment of HDACs has been observed and might potentiate the resulting hypoacetylation and decreased transcriptional activity (Vaquero et al., 2004). Conversely, H1 is associated with core histone methylations that are linked to heterochromatin formation and transcriptional repression. Specifically, H1 recruits the H3K9 histone methyltransferase Su(var)3-9 to chromatin to facilitate heterochromatin formation by heterochromatin protein-1 (HP1) (Lu et al., 2013; Healton et al., 2020). H1 is also able to recruit DNA methyltransferases to regulate imprinting regions (Fan et al., 2005; Yang et al., 2013) and may similarly be involved in regulation of transcription from CpG island promoters (Grand et al., 2021). Recruitment of modifiers depends on protein-protein interactions with the H1 CTD (Yang et al., 2013; Lu et al., 2013; Fyodorov et al., 2018).

H1 subtypes may carry out distinct functions

Evidence suggests that specific H1 subtypes may have specialized biochemical functions. Specifically, H1.2 may be involved in regulation of transcriptional elongation due to interactions with the E3 ubiquitin ligase Cul4A and elongation factor PAF1 and Pol II phosphorylated at S2 in its long unstructured C terminal domain (Kim et al., 2013). Different subtypes may also be involved in splicing of pre-mRNA. It was found that splicing factors bind to H1.0 (Kalashnikova et al., 2013b) and that H1.5 is enriched at splice sites that exhibit stalled Pol II and exon inclusion while depletion of H1.5 led to decrease in exon inclusion and a decrease in stalled Pol II (Glaich et al., 2019).

H1 compacts chromatin

Beside the biochemical functions of H1 outlined above, the most immediate and easily appreciable effect of linker histone H1 binding is the condensation of chromatin (Thoma et al., 1979) that is directly linked to transcriptional repression (Healton et al., 2020). Condensation partially depends on the high lysine content of H1 neutralizing the negative charge of DNA and alleviating electrostatic repulsion from DNA and neighboring nucleosomes (Widom, 1986; Bednar et al., 1998). In addition to the proposed charge neutralizing effect, H1 binds to nucleosomal DNA near the dyad and contacts linker DNA, drastically reducing linker DNA dynamics with effects subtly varying between H1 subtypes and contributing to nucleosome array compaction (Bednar et al., 2017; Perišić et al., 2019; Zhou et al., 2021a).

1.4 Nucleosome arrays

Compaction of nucleosome arrays is thought to modulate transcription by blocking access to DNA (Bednar et al., 1998; Klemm et al., 2019). Studies using analytical ultracentrifugation, negative stain electron microscopy, and restriction enzyme digests to probe the structure of chromatin arrays of different lengths, NRL, histone tail modifications and linker histone contents have shed much light on the biophysical nature and properties of nucleosome arrays (Dorigo et al., 2004; Robinson et al., 2006; Robinson and Rhodes, 2006; Routh et al., 2008). To understand the intricate molecular mechanisms of how structure and compaction affect chromatin function, however, molecular resolution insight into the structure of nucleosome arrays is necessary.

1.4.1 The structure of nucleosome arrays

For a long time, two models of ordered nucleosome arrangements into a chromatin fiber (30 nm fiber) were prevalent (Robinson et al., 2006; Tremethick, 2007). In the solenoid or one-start model, nucleosomes linearly lead along a helical path (Finch and Klug, 1976) whereas in the two-start or zig-zag model, the DNA path zig-zags back and forth perpendicularly to the fiber axis (Thoma et al., 1979; Woodcock et al., 1984). However, the structure and the physiological relevance of the 30 nm fiber are still controversial (Grigoryev et al., 2009; Maeshima et al., 2010; Fussner et al., 2011; Maeshima et al., 2014; Zhou et al., 2018).

Tetranucleosome structure revealed a zigzag arrangement of nucleosomes

The medium resolution crystal structure of the compacted tetranucleosome array with an NRL of 167 bp was the first structure of nucleosome arrays at molecular resolution (Schalch et al., 2005). Hereafter nucleosome arrays are referred to by the number of nucleosomes in the array followed by the NRL and the tetranucleosome array with NRL 167 bp will thus be called 4x167. The 4x167 crystal structure showed two stacks of nucleosomes connected in a zigzag way (Schalch et al., 2005) (Fig. 1.4) that is consistent with previous crosslinking studies from the same lab (Dorigo et al., 2004). The arrangement of nucleosomes is two-fold symmetric: the symmetry axis passes midway between the stacks in the linker connecting nucleosome 2 and nucleosome 3 and makes nucleosomes 1 and 2 and the connecting linker DNA symmetric with nucleosome 3 and 4 and their linker DNA (Schalch et al., 2005). In the compacted tetranucleosome, nucleosomes 3 is stacked

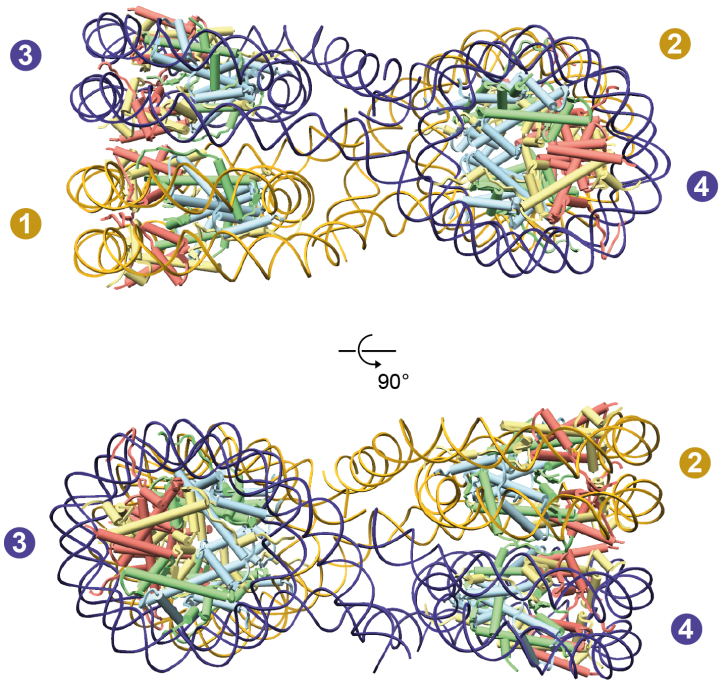

Figure 1.4: The crystal structure of the 4x167 nucleosome array. The arrangement has two-fold symmetry with nucleosomes 1 and 2 and the connecting linker DNA (gold) being symmetric to nucleosomes 3 and 4 and the connecting DNA (dark blue). In the tetranucleosome DNA zigzags back and forth between the nucleosomes of two stacks. Nucleosomes 1 and 3 form a stack and nucleosomes 2 and 4 form the second stack. The dyad axes of stacking nucleosomes are aligned and point to the opposite stack that is rotated by roughly 90°. Color code of core histones as in Fig. 1.1. PDB 1ZBB after Schalch et al. (2005).

on top of nucleosome 1 while nucleosome 4 is stacked on top of nucleosome 2 (Fig. 1.4). The two stacks are rotated almost 90° to each other, the dyad axes of stacking nucleosomes are approximately aligned and face those of the opposite stack (Schalch et al., 2005). Considering the 100 Å (10 nm) diameter of the nucleosome, the tetranucleosome is very compact as the nucleosome stack centers are only 150 Å (15 nm) apart (Schalch et al., 2005). The contact made between two stacking nucleosomes included the adjacent H2A-H2B dimers (Schalch et al., 2005) and precludes the contact of the H4 tail with the H2A-H2B acidic patch that was observed in the mononucleosome structure (Luger et al., 1997). While the structure of the tetranucleosome was informative for our understanding of nucleosome arrays, it could not provide insight into how linker histone H1 binds to and influcences the structure of nucleosome arrays.

Arrays of 12 nucleosomes containing H1 at molecular resolution

Some information on linker histone binding to arrays was won by a cryo-electron microscopy (cryo-EM) study of linker histone H1 containing arrays of 12 nucleosomes with NRL 177 bp and 187 bp (12x177 and 12x187) (Song et al., 2014). Nucleosomes zigzag

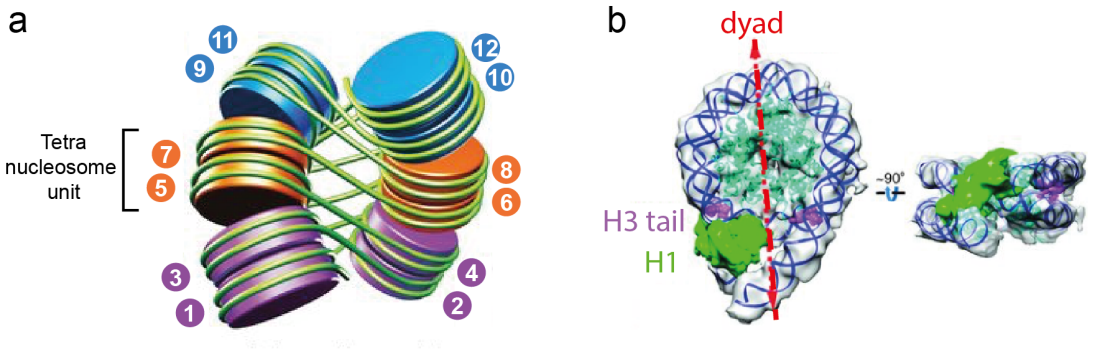

Figure 1.5: The cryo-electron microscopy structure of H1 containing 12x177 and 12x187 nucleosome arrays. **(a)** Nucleosomes are arranged in a zigzag fashion along the chromatin fiber. Tetranucleosome units (purple, orange, blue) stack on top of each other and are structurally very similar to the 4x167 crystal structure (Schalch et al., 2005). **(b)** Linker histone H1 (green) binds the nucleosomes of the array asymmetrically off-dyad. Atomic coordinates of the models for the 12x177 and 12x187 have not been deposited. Figure panels were adapted and modified from Song et al. (2014).

back and forth along the fiber axis and the $(i+2)$-th nucleosome stacks on top of the i-th nucleosome (Fig. 1.5a). While arrangement of nucleosomes was similar in the 12x177 and the 12x187, fiber dimensions in the 12x187 were larger due to the longer DNA linker. Interestingly, the structures revealed a modular composition of the fiber with tetranucleosomes similar to the 4x167 (Schalch et al., 2005) stacked on top of each in a slightly twisted way (Fig. 1.5a). The stacking between nucleosomes within the tetranucleosome unit appeared to be similar to the 4x167 with H2A-H2B dimers of stacking nucleosomes in contact, whereas stacking between tetranucleosome units was more loose (Song et al., 2014).

Linker histone H1 was observed stoichiometrically on the nucleosomes of the array (Song et al., 2014). Intriguingly, H1 localized near the dyad but was substantially shifted towards one DNA linker. In this position, the authors suggest that H1 interacts with H1 of the adjacent tetranucleosome unit to induce the twisted arrangement. The position of H1 is surprising because the structure of the nucleosome with H1.5 that has a virtually identical globular domain to the H1.4 variant used in this study showed that H1.5 binds on the dyad (Bednar et al., 2017). Additionally, H1.4 was later observed by cryo-EM to bind in the canonical on-dyad mode on the mononucleosome (Zhou et al., 2021a). This difference could be due to restraints imposed on H1 by the structure of the nucleosome array. However, a subsequent study showed that the crosslinking method used to determine the structure of the H1 bound 12x177 and 12x187 perturbs array conformation and changes how H1 interacts with the nucleosome (Zhou et al., 2018).

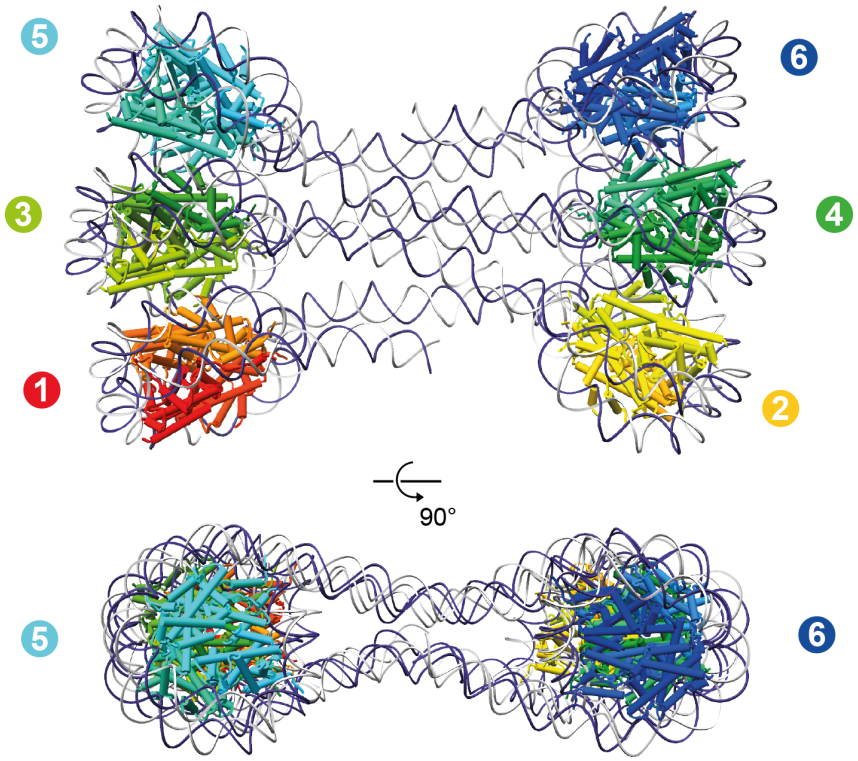

Figure 1.6: The crystal structure of an H1 containing 6x187 nucleosome array. Nucleosomes zigzag and stack loosely along the chromatin fiber to adopt a flat conformation. In solution, linker histone H1 binds to the dyad of the nucleosome but no density for H1 was observed in the crystal. PDB 6HKT (Garcia-Saez et al., 2018).

An array of 6 nucleosomes with H1

The cyrstal structure of an H1 containing array of 6 nucleosomes with NRL 187 bp (Garcia-Saez et al., 2018) showed the same zigzag arrangement of nucleosomes with the $(i+2)$-th nucleosome stacking on top of the i-th nucleosome (Fig. 1.6). Here however, nucleosomes were stacked uniformly along the fiber in a more loose way (Garcia-Saez et al., 2018) and the array adapted a flat conformation (Fig. 1.6). While localization of H1 to the dyad of the nucleosomes in the array was confirmed in solution, density for linker histone was not observed.

Binding mode of H1 to nucleosome arrays is unclear

Other medium to high resolution structures of nucleosome arrays or multi nucleosome templates with different chromatin factors exist (Ekundayo et al., 2017; Machida et al., 2018; Poepsel et al., 2018; Adhireksan et al., 2020; Zhou et al., 2021b) and have increased our understanding of how chromatin factors associate with nucleosomes in the context of longer nucleosome arrays. But despite considerable effort, the precise mode of interaction

between linker histone H1 and nucleosomal arrays remains elusive.

1.4.2 Nucleosome organization *in vivo*

In contrast to the regularly ordered chromatin arrays observed by X-ray crystallography and cryo-electron microscopy *in vitro* (Schalch et al., 2005; Song et al., 2014; Ekundayo et al., 2017; Garcia-Saez et al., 2018; Adhireksan et al., 2020; Zhou et al., 2021b), chromatin *in vivo* appears to be very heterogeneous. *In vitro* experiments and molecular dynamics simulations suggest that nucleosome arrays can adapt a wide variety of conformations (Zhou et al., 2018; Mauney et al., 2021; Ding et al., 2021; Zhurkin and Norouzi, 2021) with some enriched local configurations such as stacked nucleosomes (Mauney et al., 2021; Ding et al., 2021). Indeed, while there is evidence for regularly folded chromatin fibers in terminally differentiated chicken erythrocyte nuclei (Scheffer et al., 2011), most chromatin *in* or *ex vivo* seems to have no apparent order (Eltsov et al., 2008; Nishino et al., 2012; Chen et al., 2016; Ou et al., 2017; Cai et al., 2018; Xu et al., 2021; Beel et al., 2021). Further, nucleosomes seem to be arranged in small clusters rather than large fibers with DNA mostly zigzaging back and forth to form loosely defined tetranucleosome folding motifs (Ricci et al., 2015; Risca et al., 2017; Ohno et al., 2019; Beel et al., 2021).

1.5 Aim

Recent years have seen the development of many approaches to study chromatin organization and its impact on transcription regulation. Methods based on chromosome conformation capture (3C) technology are able to measure interaction frequencies between genomic segments and have revolutionized how we think about the genome three dimensionally. Understanding the molecular mechanisms governing chromatin processes however requires molecular resolution insight into the interactions between chromatin and chromatin binding factors.

The structures of single nucleosomes are informative but *in vivo* chromatin exists as regular arrays of nucleosomes. Structural studies of arrays in the presence or absence of linker histone H1 have provided answers for many fundamental questions about how DNA is packaged and the impact on chromatin linked processes. But studies could not confirm the existence of long chromatin fibers or regular ordered fibers *in vivo*. Additionally, the nucleosome repeat length (NRL) is linked to the transcriptional output of the underlying genomic region, and a correlation between linker histone H1 content and the NRL has been established. Do short nucleosome arrays of different NRL share common structural features? How do they change with increasing NRL? How does H1 bind to the nucleosome array? And given the link between NRL and transcriptional activity, does H1 binding depend on the NRL?

In this study we sought to determine the structures of H1 containing nucleosome arrays with different NRL to study their structure, elucidate how H1 binds to the nucleosome array and to dissect the contribution of NRL to H1 binding. For this, we reconstituted short arrays of four nucleosomes with NRL characteristic of that found near active promoter regions (4x177), in the gene bodies of actively transcribed genes (4x187 and 4x197) and in silent heterochromatin regions (4x207). We saturated the arrays with H1 and determined their structures by cryo-electron microscopy.

Materials

2.1 Chemicals and consumables

Common chemicals such as NaCl and HEPES were procured from Merck, Sigma-Aldrich, Roth and Thermo Fisher Scientific and were used interchangeably if they were of the same grade. Enzymes such as Phusion polymerase, restriction enzymes and micrococcal nuclease were generally procured from New England BioLabs except for Ulp1. Deviations from this will be stated in the methods text if applicable.

Table 2.1: Chemicals and consumables used in this work.

Chemicals/Consumables	**Manufacturer**
AG 501-X8 resin	Bio-Rad
Amicon Ultra centrifugal filters	Merck
HiTrap Q column	GE Healthcare
HiTrap S column	GE Healthcare
Instant Blue	Expedeon Inc., San Diego, US
NucleoBond PC 10000	Macherey-Nagel
NuPAGE 4 to 12%, Bis-Tris Mini Protein Gel	Thermo Fisher Scientific
NuPAGE MES SDS Running buffer	Thermo Fisher Scientific
NuPAGE MOPS SDS Running buffer	Thermo Fisher Scientific
QIAquick Gel Extraction Kit	Qiagen
QIAquick Spin Miniprep Kit	Qiagen
Slide-A-Lyzer Dialysis buttons	Thermo Fischer Scientific
SnakeSkin Dialysis Tubing	Thermo Fischer Scientific

2.2 Buffers, media and supplements

Table 2.2: Standard buffers used in this work.

Buffer	**Composition**
NuPAGE MES SDS buffer	50 mM MES, 50 mM Tris, 0.1% SDS, 1 mM EDTA, pH 7.3
NuPAGE MOPS	50 mM MOPS, 50 mM Tris, 0.1% SDS, 1 mM EDTA, pH 7.7

SDS buffer	
PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_24
Protease inhibitor cocktail	0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
TBE	89 mM Tris, 89 mM boric acid, 2 mM EDTA
TAE	40 mM Tris, 20 mM acetic acid, 1 mM EDTA

Table 2.3: Media used in this work.

Media	**Composition**
LB	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl

Table 2.4: Supplements and antibiotics used in this work.

Supplement	**Stock concentration**	**Final concentration**
Ampicillin	50 mg/mL	50 μg/mL
Chloramphenicol	34 mg/mL	34 μg/mL
Kanamycin	100 mg/mL	100 μg/mL
IPTG	0.5 mol/L	0.5 mmol/L

2.3 Bacterial strains

Table 2.5: Bacterial strains used in this work.

Strain	Genotype	Supplier
E. coli		
BL21 (DE3) RIL	*F- ompT hsdS(r_{B-} m_{B-})dcm$^+$ Tet^r gal λ(DE3) endA Hte [argU ileY leuW Cam^r]*	Agilent
BL21 (DE3) Rosetta 2 pLysS	*F- ompT gal dcm lon hsdS$_B$($r_{B-}m_{B-}$) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB$^+$]$_{K-12}$(λ^S) pLysSRARE[T7p20 ileX argU thrU tyrU glyT thrT argW metT leuW proL ori$_{p15A}$](Cam^R)*	Novagen
XL1 blue	*endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K- m_K+)*	Stratagene

2.4 Vectors

Table 2.6: Vectors used in this work.

Plasmid	Source	Resistance	Expression host	Addgene number
LIC 1B	Scott Gradia QB3 MacroLab	Kan	*E. coli* T7	29653

2.5 DNA sequences

4x177 sequence

ATCCCGGATCCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTA
GACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAA
CCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCACATATATACATC
CTGTTCCAGTGCCGGACCCGAGCATCCGGATCCCCTGGAGAATCCCGGTGCC
GAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC
GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGG
CACGTGTCACATATATACATCCTGTTCCAGTGCCGGACCCGAGCATCCGGAT
CCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCT

AGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGG
GGATTACTCCCTAGTCTCCAGGCACGTGTCACATATATACATCCTGTTCCAG
TGCCGGACCCGAGCATCCGGATCCCCTGGAGAATCCCGGTGCCGAGGCCGC
TCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTC
CCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTC
ACATATATACATCCTGTTCCAGTGCCGAT

4x187 sequence

ATCTCTCGCGCACTGGCCGCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATT
GGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGC
GTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATAT
ATACATCCTGTCATGTAAGTATTAAGGTAACCCGTCTCGCGCACTGGCCGCC
TGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCA
CCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGAT
TACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGTCATGTAAGT
ATTAAGGTAACCCGTCTCGCGCACTGGCCGCCTGGAGAATCCCGGTGCCGA
GGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGC
GCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCA
CGTGTCAGATATATACATCCTGTCATGTAAGTATTAAGGTAACCCGTCTCGC
GCACTGGCCGCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAG
ACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAAC
CGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCC
TGTCATGTAAGTATTAAGGTGAT

4x197 sequence

ATCGTCTCGCGCACTGGCCGCCATACTGGAGAATCCCGGTGCCGAGGCCGC
TCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTC
CCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTC
AGATATATACATCCTGTCATGTAAGTATTAAGGTAACCCAGTACTGTCTCGC
GCACTGGCCGCCATACTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTC
GTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTT
TAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATAC

ATCCTGTCATGTAAGTATTAAGGTAACCCAGTACTGTCTCGCGCACTGGCCG
CCATACTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCT
CTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAA
GGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGTCAT
GTAAGTATTAAGGTAACCCAGTACTGTCTCGCGCACTGGCCGCCATACTGGA
GAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGC
TTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACT
CCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGTCATGTAAGTATTA
AGGTAACCCGAT

4x207 sequence

ATCCTGGCCGCCACTGGCCGCCACTGGCCACTGGAGAATCCCGGTGCCGAG
GCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCG
CTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCAC
GTGTCACATATATACATCCTGTGCATGTAAGTGCATGTAAGTGCATGTAAGT
ACTCTGGCCGCCACTGGCCGCCACTGGCCACTGGAGAATCCCGGTGCCGAG
GCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCG
CTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCAC
GTGTCACATATATACATCCTGTGCATGTAAGTGCATGTAAGTGCATGTAAGT
ACTCTGGCCGCCACTGGCCGCCACTGGCCACTGGAGAATCCCGGTGCCGAG
GCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCG
CTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCAC
GTGTCACATATATACATCCTGTGCATGTAAGTGCATGTAAGTGCATGTAAGT
ACTCTGGCCGCCACTGGCCGCCACTGGCCACTGGAGAATCCCGGTGCCGAG
GCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCG
CTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCAC
GTGTCACATATATACATCCTGTGCATGTAAGTGCATGTAAGTGCATGTAGAT

Table 2.7: DNA sequences used in this work.

Name	sequence
v1fw Smt3	TACTTCCAATCCAATGCATCGGACTCAGAAGTCAATCAAGAAGCTAAGC
Smt3 rv	CCCACCAATCTGTTCTCTGTGAGCC
Smt3-H1.4	CTCACAGAGAACAGATTGGTGGGAGCGAAACCGCACCGG
H1.4-GyrA	ACCAGTGCATCACCGGTAATACATTTCTTTTTGGCTGCTGCCTTCTTAGG
GyrA fw	TGTATTACCGGTGATGCACTGGTTGC
v1rv GyrA	TTATCCACTTCCAATGTTATTAATGGTGATGATGATGGTGATGGGTTG

2.6 Recombinant plasmids

Table 2.8: Recombinant plasmids used for protein expressions in this work. Asterisks denote full length protein sequence. All sequences for histones are human, Ulp1 sequence is from *Saccharomyces cerevisiae*. Linker histone H1.4 sequence after cleavage described in Methods is the full length protein without leading methionine.

Factor	Residues	Tag(s)	Vector	Selection	Source
H2A type 1-B/E	1-130 *	N-His_6	LIC 1B	Kan	Dodonova et al., 2020
H2B type 1-K	1-126 *		pET22b	Amp	Dodonova et al., 2020
H3.2	1-136 *		pET22b	Amp	Dodonova et al., 2020
H4	1-103 *		pET3a	Amp	Dodonova et al., 2020
H1.4	2-218	N-His_6-Smt3 GyrA-His_6-C	LIC 1B	Kan	this study
Ulp1	403-621	N-His_6	pFGET19	Kan	Addgene 64697

Table 2.9: Recombinant plasmids used for DNA template production in this work. DNA templates were designed based on the tetranucleosomal unit in the 12x177 (Song et al., 2014) and ordered from GeneArt (Thermo Fisher Scientific).

Template	**Length (bp)**	**Vector**	**Selection**	**Source**
4x177	702	pMX	Kan	GeneArt
4x187	748	pMX	Kan	GeneArt
4x197	788	pMX	Kan	GeneArt
4x207	828	pMX	Kan	GeneArt

Methods

3.1 Gel electrophoresis

3.1.1 Agarose gel electrophoresis

Products of restriction enzyme digests, PCR, micrococcal nuclease digest and PEG precipitation were analyzed by agarose gel electrophoresis on a gel containing 1% (w/v) agarose in TAE buffer with SYBR safe DNA gel stain (Invitrogen) unless stated otherwise. Samples were mixed with 6x DNA Loading Dye (New England BioLabs) to a final concentration of 1x dye and loaded onto the gel. For comparison, the 1 kb DNA ladder (New England BioLabs) was loaded onto the same gel. The gel was run at 120 V for 20 to 30 min until separation of samples was satisfactory. Visualization of DNA was done using the INTAS GEL iX20 Imager system.

3.1.2 Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Separation of proteins according to their molecular weight was performed by sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) during protein purification. Samples were mixed with 4x loading dye (250 mM Tris-HCl pH 6.8, 10% SDS, 0.01 % bromophenol blue, 20% β-mercapto ethanol, 40% glycerol) to a final concentration of 1x dye and incubated at 95 °C for 5 min and loaded onto a NuPAGE 4-12% gradient Bis-Tris Protein gel (Invitrogen). For comparison, the PageRuler Prestained Protein Ladder (Thermo Fisher Scientific) was loaded as molecular weight standard. Electrophoresis was done in 1x MOPS or 1x MES SDS-PAGE buffer (Invitrogen) at 200 V for 45 min. Protein was visualized by incubating with Instant Blue (Invitrogen) for 1 h or over night and destaining by incubating with deionized water. Gels adequately stained and destained were imaged using an Epson Perfection V800 flatbed scanner.

3.1.3 Electrophoretic mobility shift assay (EMSA)

Samples from nucleosome array reconstitutions with or without linker histone H1.4 were analyzed by electrophoretic mobility shift assay (EMSA). Before loading, the sample was adjusted to 300 ng DNA in 14 μL of HEN 0 buffer (10 mM HEPES pH 7.0, 1 mM DTT, 1 mM EDTA) and 1 μL of HEN 0 containing 30% glycerol was added for a total sample volume of 15 μL. Sample was analyzed on a 1.2% agarose in 0.5x TBE buffer for 3 h at 100 V at 4 °C using the 1 kb DNA ladder (New England BioLabs) as a size standard. After electrophoresis, the gel was incubated in 0.5x TBE buffer containing SYBR safe DNA gel stain (Invitrogen) for 30 min and imaged using the INTAS GEL iX20 Imager system.

3.2 DNA methods

3.2.1 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify DNA sequences and to fuse gene fragments by using overlapping primers. Phusion High-Fidelity DNA polymerase (New England BioLabs) was used according to the manufacturer's instructions in a 50 μL reaction containing 0.5 μM forward primer, 0.5 μM reverse primer, 50 ng DNA template, 200 μM dNTPs, 1x Phusion HF buffer (New England BioLabs) and 1 unit Phusion DNA polymerase. A standard PCR program (Table 3.1) was run on a Professional TRIO Thermocycler (Analytik Jena), the product was analyzed by agarose gel electrophoresis and the band corresponding to the desired product was extracted using the Qiagen QIAquick Gel extraction kit according to the manufacturer's instructions.

3.2.2 Transformation of chemically competent *E. coli*

Frozen chemically competent *E. coli* suspensions were thawed on ice and added to plasmid DNA solutions. The mixture was incubated on ice for 30 min, then heat shocked at 42 °C for 45 s and kept on ice for 2 min. After that, 1 mL of LB medium was added and cells were incubated for 1 h at 37 °C and 500 rpm in an Eppendorf Thermomixed. Cells were plated on LB agar plates containing the appropriate selection antibiotic and incubated over night at 37 °C. Material from plates was used directly or plates were stored at 4 °C

Table 3.1: Polymerase chain reaction (PCR) protocol for the amplification and fusion of DNA sequences. Annealing temperature and elongation time were modified to fit primer pairs and length of product, respectively.

Step	Temperature	Time	Repeat
1. Initial denaturation	98 °C	30 s	
2. Denaturation	98 °C	15 s	
3. Annealing	60-70 °C	30s s	
4. Elongation	72 °C	30 s / kb s	go to step (2) 30x
5. Final elongation	72 °C	10 min	
6. Hold	4 °C	hold	

until use.

3.2.3 Ligation independent cloning (LIC)

PCR

DNA sequences coding for Smt3 (plasmid coding for an Smt3-tagged protein was a gift from S. Schilbach), human H1.4 (synthesized by IDT and codon optimized for *E. coli*), and GyrA (synthesized by IDT) were amplified by PCR. Products were extracted from an agarose gel and fused by overlap PCR using primers that also introduced the LIC adapters.

Plasmid linearization

Restriction enzyme digests were used to linearize the target vector for insertion of DNA prepared by overlap PCR. For this, 0.6 μg of vector LIC-1B were mixed with 5 μL of 10x CutSmart Buffer (New England BioLabs), 1 μL of the enzyme SspI (New England BioLabs), adjusted to a total volume of 50 μL, and incubated at 37 °C for 1 h. The reaction was analyzed by agarose gel electrophoresis and the linearized vector was extracted using the Qiagen QIAquick Gel extraction kit according to the manufacturer's instructions.

LIC reaction

Insert prepared by overlap PCR and linearized target plasmid were treated with T4 DNA polymerase (New England BioLabs) in the presence of complementary dNTP to generate complementary overhangs. For this, 10 μL of gel purified vector or 10 μL of gel purified PCR product were mixed with 2 μL of dCTP (Thermo Fisher Scientific) for LIC-1B or

dGTP (Thermo Fisher Scientific) for PCR product, 1 μL of 100 mM DTT, 2 μL of 10x T4 DNA pol buffer (New England BioLabs), 4.6 μL of deionized water and 0.4 μL of T4 DNA polymerase for a total reaction volume of 20 μL. The reactions were incubated at 22 °C for 30 min and then at 75 °C for 20 min. For transformation into chemically competent *E. coli* XL1 blue, 2μL of the vector LIC reaction and 2 μL of the PCR product LIC reaction were mixed with 6 μL of deionized water, incubated at room temperature for 10 min, and then transformed into 30 μL of competent cells.

3.2.4 Preparation of DNA for nucleosome reconstitution

Amplifying DNA in *E. coli* XL1 blue cells

Plasmids containing the four repeats of the strong nucleosome positioning sequence Widom-601 (Lowary and Widom, 1998) flanked by EcoRI restriction enzyme sites were transformed into *E. coli*, plated onto LB agar plates containing kanamycin, and incubated over night at 37 °C. The next morning, material from the plates was used to inoculate 25 ml of LB supplemented with kanamycin and incubated at 37 °C and 150 rpm for 8 h. Then, 6 large unbaffled flasks each with 1L of LB supplemented with kanamycin were inoculated 1:1,000 and incubated at 37 °C and 150 for 16 h. The absorbance per mL of the cell suspension at 600 nm (OD_{600}) was measured, suspension was centrifuged for 20 min at 6,000 rpm in a Sorvall LYNX 6000 centrifuge (Thermo Scientific) using an F9 rotor. Pellets equivalent to a total OD_{600} of 6,000 were pooled and used for plasmid preparation using a NucleoBond PC 10000 kit (Macherey-Nagel) according to the manufacturer's instructions or frozen in liquid nitrogen and stored at -80 °C until use.

DNA fractionation by polyethylene glycol (PEG) precipitation

To separate insert from plasmid backbone, 10 mg of plasmid were used in a 5 mL reaction containing 100 μL of EcoRI (New England BioLabs) and 500 μL 10x CutSmart Buffer (New England BioLabs) and incubated over night at 37 °C. The sample was adjusted to 800 mM NaCl by adding 5 M NaCl stock solution and incubated at 37 °C for 20 min. Then, the sample was adjusted to 5% PEG-6000 by adding 40% polyethylene glycol (PEG)-6000 solution and incubated for 10 min at room temperature. The sample was divided into 2 mL Eppendorf reaction tubes, spun down for 10 min at 13.200 rpm in an Eppendorf 5424R table top centrifuge and the supernatant was transferred into a fresh 15 mL Falcon. The pellets in the 2 mL Eppendorf tube were washed with 200 μL 70%

ethanol and air dried. PEG-6000 concentration in the supern was increased by 0.5% and the previously described steps were repeated. The last fractionation step was at 7.5% PEG-6000. The air dried pellets were resuspended in 50 μL deionized water. Digest and fractionation were monitored by agarose gel electrophoresis and fractions containing the DNA insert were pooled and stored at -20 °C until use.

3.3 Protein purification

Unless stated otherwise, all purification steps were performed at 4 °C. For sonification, a Branson sonifier was used and for centrifugation steps, a Sorvall LYNX 6000 centrifuge (Thermo Scientific) with an F9 rotor was used. Chromatography columns were pre-equilibrated with the buffer the sample is contained in and maintained and cleaned according to manufacturer's instructions and reused for the same samples. Protein expression and purification were monitored by SDS-PAGE.

3.3.1 Ulp1 purification

Ulp1 was expressed from pFGET19-Ulp1 (a gift from Hideo Iwai, Addgene plasmid #64697) and purified as previously reported (Guerrero et al., 2015).

3.3.2 Human core histone purification

Human core histones were expressed and purified from inclusion bodies as reported previously for the more widely used *Xenopus laevis* core histones (Luger et al., 1999; Dyer et al., 2004).

Core histone expression

Plasmids encoding human H2B, H3 and H4 were a gift from the W. Fischle lab, plasmid encoding His-tagged H2A was a gift from S. Dodonova. Chemically competent *E. coli* BL21 (DE3) RIL cells were transformed, used to inoculate 25 ml LB selection medium, and incubated over night at 37 °C and 150 rpm. Then, 6 baffled flasks containing 1 L of LB selection medium were inoculated 1:1,000 and incubated at 37 °C and 150 rpm until reaching OD_{600} of 0.4-0.8 / mL. Cell suspension was adjusted to 0.5 mM IPTG and incubated at 37 °C and 150 rpm for 4 h and centrifuged for 20 min at 6,000 rpm. Pellets were used for purification or frozen in liquid nitrogen and stored at -80 °C until use.

Inclusion body purification and extraction

Pellets from 6 L were resuspended in 100 mL lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT, 5% glycerol, 1x protease inhibitor cocktail) and lysed by sonication for 10 min with 10 s on pulses and 20 s off pulses at 30%. The lysate was spun down for 20 min at 27,000 xg at 4 °C and the pellet was washed once with TW buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 1x protease inhibitor cocktail, 1% Triton X-100) and twice with wash buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 1x protease inhibitor cocktail). For the washes, the pellet was resuspended in buffer using a manual Douncer (Sigma Aldrich) and spun down for 20 min at 18,000 xg at 4 °C.
The pellet was resuspended in 2 mL DMSO and incubated at room temperature for 30 min. Then, 30 ml of unfolding buffer (7M guanidinium hydrochloride, 20 mM HEPES pH 7.5, 10 mM DTT) were added and the pellet was resuspended further by Douncing and incubated at room temperature for 1 h on a magnetic stirrer. The sample was spun down for 20 min at 23,000 xg at 4 °C, the supernatant was collected and kept on ice, and the pellet was again Dounced in 10 mL unfolding buffer, stirred for 30 min at room temperature, and spun down 20 min at 23,000 xg at 4 °C. The supernatant was pooled and dialized against 2 L of SAU-200 (7M urea deionized for 1 h with AG-501 X8 resin, 20 mM NaOAc pH 5.2, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) for 3 h followed by 2 L of SAU-200 over night using SnakeSkin dialysis tubing (Thermo Fisher Scientific).

Cation exchange chromatography

The dialysate was cleared by centrifugation for 20 min at 27,000 xg at 4 °C and purified by running it over a HiTrap SP 5 mL HP column (GE Healthcare) followed by a HiTrap Sp 5 mL HP column (GE Healthcare) using an ÄKTA pure system (GE Healthcare). After loading the sample, the Q column was disconnected and the sample was eluted off the Sp column using a linear gradient of SAU-1000 (7M urea deionized for 1 h with AG-501 X8 resin, 20 mM NaOAc pH 5.2, 1000 mM NaCl, 1 mM EDTA, 1 mM DTT) in SAU-200 over 20 column volumes. Peak fractions containing core histones were pooled and dialyzed twice against 2 L of dialysis buffer (20 mM HEPES pH 7.5, 1 mM DTT). Aliquots of 3 mg were lyophilized (Alpha 1-2 LDplus, Christ) and stored at -80 ° until use.

3.3.3 Human linker histone purification

Human linker histone H1.4 was expressed and purified as reported previously (Osunsade et al., 2019) with modifications.

Smt3-H1.4-GyrA expression

Plasmid encoding human Smt3-H1.4-GyrA was transformed into chemically competent *E. coli* BL21 (DE3) Rosetta 2 pLysS cells. Transformed cells were used to inoculate 25 ml LB selection medium and incubated over night at 37 °C and 150 rpm. Then, 6 baffled flasks containing 1 L of LB selection medium were inoculated 1:1,000 and incubated at 37 °C and 150 rpm until reaching OD_{600} of 0.3 / mL. Cells were incubated at 16 °C and 150 rpm for 30 min, adjusted to 0.5 mM IPTG and incubated at 16 °C and 150 rpm overnight. The cell suspension was centrifuged for 20 min at 6,000 rpm. Pellets were used for purification or frozen in liquid nitrogen and stored at -80 °C until use.

Purification of Smt3-H1.4-GyrA and tag cleavage

Pellets from 6 L were resuspended in 50 mL lysis buffer (100 mM HEPES pH 7.5, 600 mM NaCl, 1x protease inhibitor cocktail) and lysed by sonication for 10 min with 5 s on pulses and 10 s off pulses at 20%. The lysate was spun down for 10 min at 15,000 rpm at 4 °C and the supernatant was loaded onto a HisTrap HP 5 mL column (GE Healthcare) using an ÄKTA pure system (GE Healthcare). The column was washed twice with 9 column volumes of wash buffer (20 mM HEPES pH 7.5, 2000 mM NaCl, 60 mM imidazol) followed by 5 column volumes of lysis buffer and eluted with a gradient of elution buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 500 mM imidazol; pH adjusted to 7.5) in lysis buffer over 15 column volumes. Peak fractions containing full length Smt3-H1.4-GyrA were pooled, adjusted to 1 mM DTT, supplemented with 1 mg of Ulp1 and incubated at room temperature for 1 h. Then, sample was adjusted to 500 mM β-mercapto ethanol, incubated at room temperature for 4 h, and adjusted to 8 M urea by weighing in solid urea.

Purification of H1.4

The sample was then added to 1 L of buffer A (50 mM Tris-HCl pH 9.0), loaded onto a HiTrap Sp 1 mL HP column (GE Healthcare) and eluted overnight with a shallow gradient of buffer B (50 mM Tris-HCl pH 9.0, 1000 mM NaCl). Then, 1/10 volume of 1

M HEPES pH 7.0 was added and the sample was run over a HisTrap HP 1 ml column (GE Healthcare). The flowthrough was collected and dialyzed twice against 1 L of dialysis buffer (20 mM HEPES pH 7.0, 600 mM NaCl) for 3 h and concentrated using Amicon 0.5 mL 10 kDa MWCO centrifugal filters. The concentration was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions and the protein was aliquoted, frozen in liquid nitrogen and stored at -80 °C until use.

3.4 Nucleosome reconstitution

3.4.1 Histone octamer reconstitution

Histone octamers from human core histones were reconstituted as previously described (Dyer et al., 2004) with minor modifications. Lyophilized human core histone aliquots were resuspended in 1mL unfolding buffer (7M guanidinium hydrochloride, 20 mM HEPES pH 7.5, 10 mM DTT) and incubated on a rotating wheel at 12 rpm for 30 min at 4 °C. H2A, H2B, H3 and H4 were mixed at molar ratio of 1.2:1.2:1:1 and dialyzed three times against 2 L of refolding buffer (10 mM HEPES pH 7.5, 1 mM EDTA, 2 M NaCl, 2.5 mM DTT) for 3 h, over night, and 3 h, respectively, in a SnakeSkin 7 kDa MWCO dialysis tubing. The dialysate was recovered, cleared by spinning down in an Eppendorf 5424R table top centrifuge for 10 min at top speed at 4 °C, concentrated to a volume < 500 μL using Amicon 0.5 mL 10 kDa MWCO centrifugal filters, and applied to a pre-equilibrated Superdex 200 Increase 10/300 size exclusion column. Peak fractions containing intact core histone octamer were pooled, aliquoted, frozen in liquid nitrogen and stored at -80 °C until use.

3.4.2 Nucleosome array reconstitution

Nucleosome arrays containing linker histone H1.4 were reconstituted by salt-gradient dialysis as previously described (Song et al., 2014).

Core histone octamer titration

Human core histone octamer was mixed with template DNA at different molar ratios of octamer:nucleosome positioning sequences at around 1 in buffer HEN 2.0 (10 mM HEPES pH 7.0, 2000 mM NaCl, 1 mM EDTA, 1 mM DTT) and transferred into Slide-A-Lyzer

3.5 kDa MWCO dialysis buttons. The sample was transferred into a beaker containing 500 mL of HEN 2.0 buffer. A peristaltic pump set to a flow rate of 1.83 mL/min was used to continuously add 2 L HEN 0 (10 mM HEPES pH 7.0, 1 mM EDTA, 1 mM DTT) to the beaker while subtracting excess buffer from the beaker for 16 h. The sample was recovered and spun down in a table top centrifuge at top speed and 4 °C for 10 min.
To probe octamer binding, sample containing 400 ng DNA was digested with 0.5 μL BanI (New England BioLabs), 2 μL 10x CutSmart Buffer in a total reaction volume of 20 μL for 1 h at 37 ° and analyzed by agarose gel electrophoresis.

Linker histone titration

Human core histone octamer was mixed with template DNA at the molar ratio determined by core histone octamer titration and reconstituted as described above by salt-gradient dialysis from buffer HEN 2.0 to buffer HEN 0.6 (10 mM HEPES pH 7.0, 600 mM NaCl, 1 mM EDTA, 1 mM DTT). Sample was recovered and human linker histone H1.4 was added at molar ratios of H1:nucleosome positioning sequences at around 1. The sample was then dialyzed further for 6 h using a flow rate of 6.67 mL/min to gradually dialyze to 2 L HEN 0. The sample was recovered and spun down in a table top centrifuge at top speed and 4 °C for 10 min. To probe linker histone binding, sample containing 300 ng DNA was analyzed by electrophoretic mobility shift assay (EMSA).

3.5 Single particle cryo-electron microscopy

3.5.1 Sample preparation for cryo-electron microscopy

QuantiFoil Cu 300 R1.2/1.3 grids were glow-discharged for 100 s at 15 mM and 0.4 mbar using a PELCO easiGlow System (Ted Pella). In a Vitrobot Mark IV (FEI Company) with chamber humidity set to 100%, temperature set to 16 °C and blotting papers equilibrated for 1 h, 2 μL of H1 containing tetranucleosome sample at 100 ng DNA/μL were applied from each side[†]. Excess liquid was blotted away using blot force 5 for 3 s and the grid was vitrified by plunging into liquid ethane.

[†]Sample preparation for the 4x177 sample was performed by M. Engeholm during SARS-CoV-2 induced shut-down of the institute.

3.5.2 *Cryo-EM data collection

Data were collected† on a Titan Krios 300 kV transmission electron microscope (FEI Company) equipped with a Gatan Imaging Filter set to 20 eV and a K3 direct electron detector (Gatan). Movies containing 60 frames with a total fluence of 60 $e^-/Å^2$ were collected using SerialEM at a nominal magnification of 81,000x and a pixel size of 1.05 Å/pixel with 40° stage tilt.

†Data collection for the 4x177 sample was performed by C. Dienemann during SARS-CoV-2 induced shut-down of the institute.

3.5.3 *Cryo-EM data processing

Gain normalization, motion correction, and CTF estimation of cryo-EM movies were performed using Warp and particles were picked using an instance of Warp?s neural network retrained on the 4x177 data set. Particles were extracted at 8.4 Å/pixel in RELION 3.1 and sorted by 2-3 rounds of 2D classification in cryoSPARC. Particles belonging to classes showing two or more nucleosomes were reextracted at 3.15 Å/pixel and all subsequent processing was done in RELION 3.1.

For the 4x177+H1.4 data set†, several rounds of 3D classification yielded particles that were refined to a 7.2 Å resolution map of a 4x177 trinucleosome. From this, 3D classification with a mask around the presumed location of the nucleosome 4 yielded particles that were refined to a 9.5 Å resolution map of the 4x177 tetranucleosome. The signal of the trinucleosome was subtracted from these particles and the output was refined to the 7.9 Å resolution map of the fourth nucleosome. From the 4x177 trinucleosome map, masked refinements on the nucleosome stack or the connecting nucleosome, respectively, were signal subtracted for the other nucleosomes and refined to yield the focused refined maps of nucleosome 1, 2 and 3, respectively.

The 4x187, 4x197 and 4x207 cryo-EM data were subjected to several rounds of 3D classification and 3D refinement to yield maps with a defined nucleosome stack and blurred density for the connecting nucleosome. From this map, several more rounds of 3D classification were performed and the selected particles were refined to the 4x187, 4x197 and 4x207 trinucleosome at 11 Å, 9.7 Å and 9.8 Å resolution, respectively. Particles from the 3D refinement of the stack with less defined connecting nucleosome were extracted, unbinned and further processed using signal subtraction, 3D classifications, and masked refinements to yield maps for nucleosome 1, 2 and 3. For the 4x187 dataset the same

strategy was applied to obtain the map for nucleosome 4 but proved unsuccessful for the 4x197+H1.4 and 4x207+H1.4 datasets. The angular distribution of views for each map was plotted using Warp, local resolution and global FSC was determined using RELION, and the directional FSCs were calculated using the 3D FSC server.

[†]Image processing for the 4x177 sample was performed by S. Dodonova during SARS-CoV-2 induced shut-down of the institute.

3.5.4 *Structural model building

The local resolution-filtered maps were used for model building except for the 4x177 trinucleosome, 4x177 nucleosome 1, 4x177 nucleosome 2 and 4x177 nucleosome 4 for which the post-processed maps were used. For each data set, the structure of the H1-bound mononucleosome (PDB ID 7K5Y) with protein and DNA sequences mutated to the ones used in this study was rigid-body fitted into the density of nucleosomal unit in UCSF chimera. Protein termini, entry DNA and exit DNA were manually adjusted in COOT and the resulting structures were real-space refined in PHENIX. The refined nucleosomal units were then rigid-body fitted into corresponding densities of the nucleosome stack, trinucleosome and tetranucleosome, respectively, using UCSF Chimera. In case of the trinucleosome and tetranucleosome structures, the linker DNA was manually built in COOT. The models were real-space refined in PHENIX and validated using Molprobity (Tables X). Figures were generated using PyMOL (Schrödinger LLC), UCSF Chimera, and UCSF ChimeraX.

3.5.5 *Analysis of linker DNA trajectories

The models for the nucleosome stacks were used to measure linker DNA trajectories for nucleosomes 1 and 3, and the models of the focused-refined maps of nucleosomes 2 and 4 were used to measure linker DNA deviation for nucleosomes 2 and 4. The corresponding maps were used to rigid-body fit the structure of the H1-bound 197 bp mononucleosome (PDB 7K5Y). The plane of the nucleosome disc needed to be defined to determine the angle α and a plane normal to the nucleosome disc along the dyad axis needed to be defined to determine the angle β. For definition of these planes, three points for each nucleosomal unit were defined: (1) the centroid of the coordinates of the central base pair of the 147 bp Widom-601 sequence, (2) the centroid of the coordinates of the base pair 38 bp upstream of 1 and (3) the centroid of the coordinates of the base pair 39 bp downstream

of 1. Points (2) and (3) are on two different DNA gyres and on the opposite side of the nucleosome dyad. Vector **v** was defined using points (2) and (3) to approximate the normal to the nucleosome disc and vector **u** was defined using point (1) and the centroid of points (2) and (3) to approximate the dyad axis. Vectors **u** and **v** were used to describe the plane perpendicular to the nucleosome disc. The normal **w** to this plane was defined by taking the normalized cross product of **u** and **v**. Then, vectors **u** and **w** were used to describe the plane of the nucleosome disc. Linker DNA vectors were defined by using (4) the centroid of coordinates of the base pair 5 bp into the Widom-601 sequence and (5) the centroid of the coordinates of the base pair 10 bp outside of the Widom-601 sequence. For measurement of the angle β, linker DNA vectors were projected onto the plane generated by **u** and **v** and the angle between the projected vectors was calculated. For the angle α, linker DNA vectors were projected onto plane the plane generated by **u** and **w** and the angle between the projected vectors was calculated.

Coordinates were extracted from the `.pdb` file with `bash` scripting and calculations were performed with MATLAB R2017a.

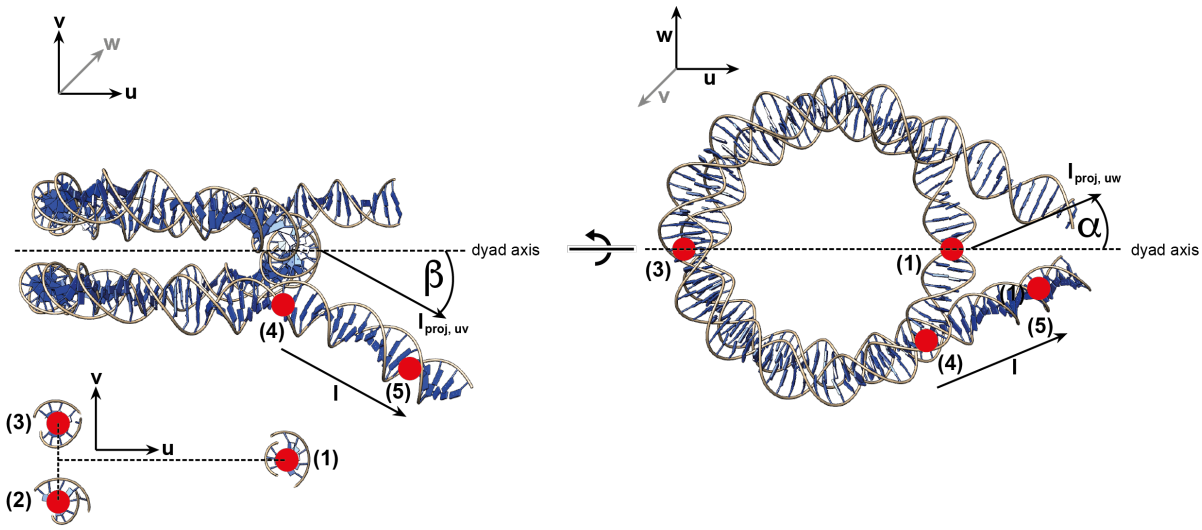

Figure 3.1: Analysis of linker DNA trajectories. The DNA model of nucleosome 3 from the 4x177 array is shown here from two views to illustrate the two different angles β (left) and α (right) of linker DNA (vector **l**) relative to the nucleosome dyad axis5. To define the plane (**u**,**v**) and (**u**,**w**) in which to measure linker DNA angle β and α, respectively, we defined points (1) the centroid of the coordinates of the central base pair of the 147 bp Widom-601 sequence, (2) the centroid of the coordinates of the base pair 38 bp upstream of 1 and (3) the centroid of the coordinates of the base pair 39 bp downstream of 1. We defined vectors **v** using points (2) and (3) and **u** using point (1) and the centroid of points (2) and (3) and determined the normal **w** to this plane by taking the normalized cross product of **u** and **v**. Linker DNA vectors were defined using (4) the centroid of coordinates of the base pair 5 bp into the nucleosome and (5) the centroid of the coordinates of the base pair 10 bp outside of the nucleosome. To measure differences in β, linker DNA vectors from nucleosomes of the different NRL arrays and the reference H1 bound nucleosome (PDB ID 7K5Y) were projected onto the plane (**u**,**v**) and the angle between the projected vectors was calculated. To measure differences in α, linker DNA vectors from nucleosomes of the different NRL arrays and the reference H1 bound nucleosome (PDB ID 7K5Y) were projected onto the plane (**u**,**w**) and the angle between the projected vectors was calculated.

Results

4.1 Reconstitution of tetranucleosome arrays

4.1.1 Core histone purification and octamer reconstitution

Purification of core histones is well established (Luger et al., 1999; Dyer et al., 2004). Human core histones H2A, H2B, H3 and H4 were expressed in *E. coli* and purified from inclusion bodies by cation exchange chromatography (Fig. 4.1). Purified histones were relatively pure and typical yields per purification ranged from 30 mg to 100 mg. Core histone octamer was reconstituted at molar ratios H2A:H2B:H3:H4 of 1.2:1.2:1:1 from lyophilized human core histones. As H2A and H2B exist as a dimer and H3 and H4 exist as a dimer and a dimer of dimers in solution, these needed to be separated from the full octamer by size exclusion chromatography (Fig. 4.2) The first large peak at elution volume 11 ml corresponds to the octamer, the H3-H4 tetramer, if present, would be visible as a shoulder to that peak and the H2A-H2B dimer separates well at elution volume 14 ml (Luger et al., 1999; Dyer et al., 2004). For reconstitution of nucleosomes, the first few fractions of the octamer peak were pooled and concentrated.

4.1.2 Linker histone H1.4 purification

As we wanted to study linker histone binding to arrays free from any conceivable artifacts, a strategy to purify full length H1.4 without any leftover amino acids from tag cleavage was adapted (Osunsade et al., 2019). For this, H1.4 was cloned with an N-terminal His-tagged Smt3 and a C-terminal His-tagged protein sequence containing the intein from *Mycobacterium xenopi* GyrA (Osunsade et al., 2019). After immobilized metal affinity chromatography via the His-tags (Fig. 4.3), the N-terminal Smt3 and the C-terminal GyrA were cleaved by Ulp1 digest and thiolysis with high concentration of β-mercapto ethanol, respectively (Fig. 4.3), yielding the final full length native H1.4. Due to non-stoichiometric cleavage of the GyrA tag, the sample had to be purified further. The sample was concentrated by cation exchange chromatography and purified by reverse affinity chromatography to sort out uncleaved protein and the cleaved tags (Fig. 4.3). Protein identity was confirmed by mass spectrometry.

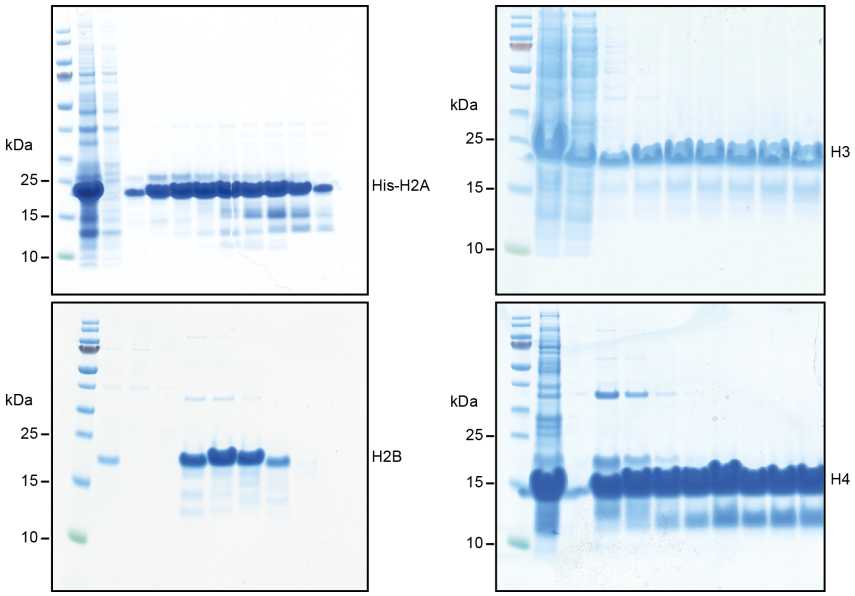

Figure 4.1: Human core histone H2A, H2B, H3 and H4 purification. Histones were expressed in *E. coli* and purified from inclusion bodies as previously described (Luger et al., 1999; Dyer et al., 2004). SDS-PAGE shows cation exchange chromatography of H2A, H2B, H3 and H4.

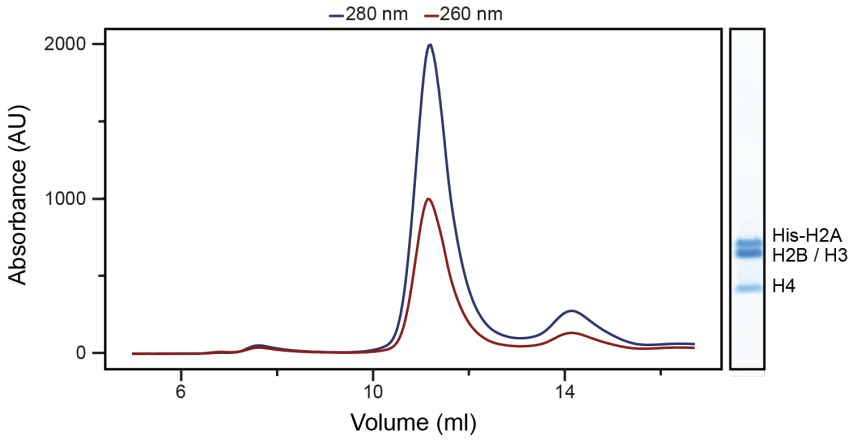

Figure 4.2: Histone octamer reconstitution from purified core histones. Core histone octamer was reconsituted from purified human core histones as previously described (Luger et al., 1999; Dyer et al., 2004). Reconstituted histone octamer was purified by size exclusion chromatography and analyzed by SDS-PAGE.

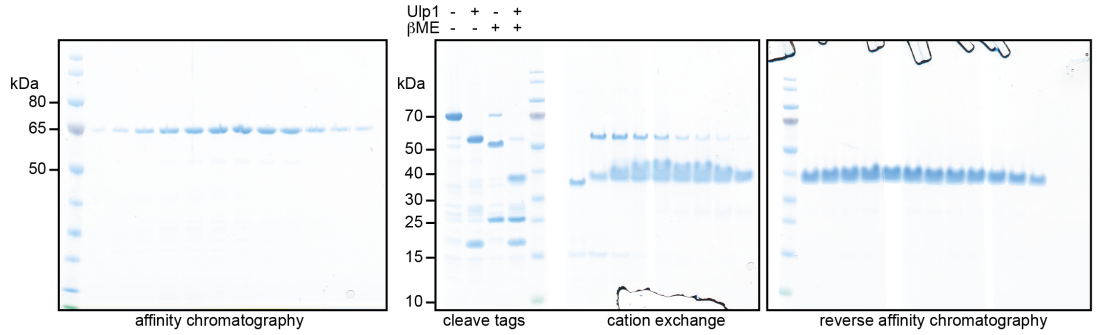

Figure 4.3: Human linker histone H1.4 purification. Human linker histone H1.4 was purified as previously described to yield full length H1 without additional residues remaining after tag cleavage (Osunsade et al., 2019). His_6-Smt3-H1.4-GyrA-His_6 was purified by affinity chromatography and the N-terminal Smt3 tag and C-terminal GyrA tag were cleaved by Ulp1 digestion and by incubation with β-mercapto ethanol, respectively. The sample was concentrated by cation exchange and purified by reverse affinity chromatography to yield full length H1. Purification steps were monitored by SDS-PAGE. Identity of digestion products and final protein were confirmed by mass spectrometry.

4.1.3 DNA template preparation

Plasmids contained the four repeats of the Widom-601 nucleosome positioning sequence comprising 147 base pairs of DNA wrapped around the core histone octamer and linker DNA connecting the nucleosome positioning sequences. Between constructs, the linker DNA varied by integer multiples of 10 bp to approximate the helical repeat of the DNA duplex. Thus, the DNA scaffolds contained 30 bp, 40 bp, 50 bp and 60 bp of DNA linking each nucleosome positioning sequence and correspond to nucleosome repeat lengths (NRL) of 177 bp, 187 bp, 197 bp and 207 bp, respectively. The shorter 4x177 correspods to linker lengths found near active promoters and enhancers, while the intermediate 4x187 and 4x197 correspond to linker lengths found in gene bodies of active genes and the longer 4x207 has linker lengths similar to those observed in heterochromatin (Valouev et al., 2011). Both the DNA scaffolds and the reconstituted nucleosome arrays will be referred to as 4x177, 4x187, 4x197 and 4x207 (Fig. 4.4). The scaffolds were flanked by EcoRI restriction enzyme recognition sites. Plasmids were amplified by large cultures of *E. coli*. The plasmid was isolated and digested with EcoRI and the insert was separated from the backbone by precipitation using polyethylene glycol 6000 (Lis and Schleif, 1975) (Figure 4.4). Fractions containing only the insert at 6.0% (4x177, 4x197, 4x207) or 6.5% (4x187) PEG-6000 were used for tetranucleosome reconstitution.

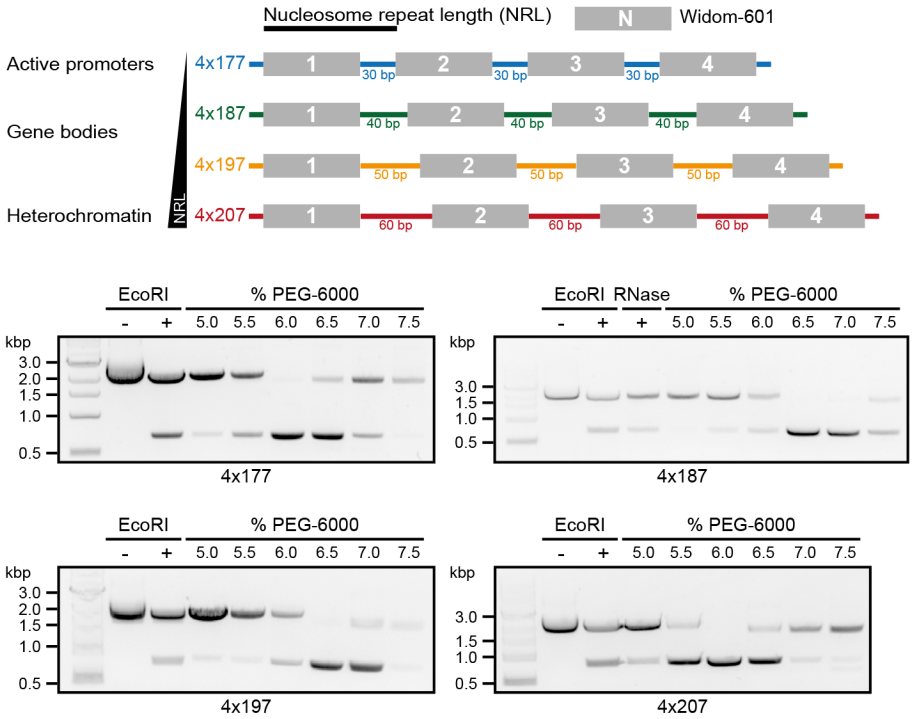

Figure 4.4: Preparation of DNA scaffolds for tetranucleosome assembly. *E. coli* transformed with plasmids carrying the tandem repeats of Widom-601 nucleosome positioning sequence flanked by EcoRI restriction enzyme sites were grown large scale and the plasmid was isolated. The insert was excised by EcoRI digest followed by an additional RNase digest for the 4x187 due to high contamination of RNA as measured using a Qubit system. Fractionation by molecular weight was performed by increasing polyethylene glycol 6000 (PEG-6000) and sedimentation (Lis and Schleif, 1975). Purification was monitored by agarose gel electrophoresis. Fractions containing only insert were used for tetranucleosome reconstitution.

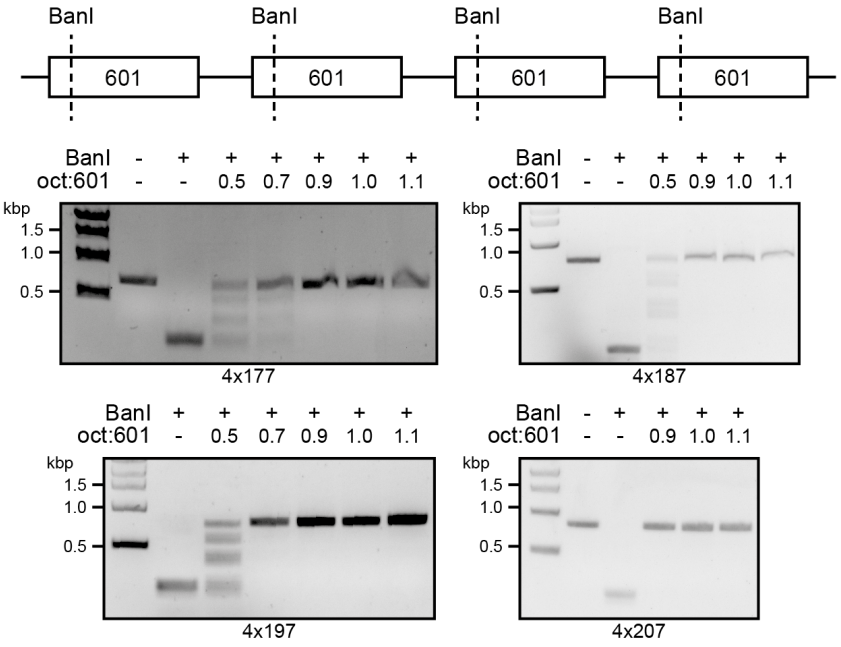

Figure 4.5: Core histone octamer titration. The Widom-601 nucleosome positioning sequence (Lowary and Widom, 1998) contains a BanI restriction enzyme site. DNA scaffolds containing four repeats of the Widom-601 sequence were reconstituted with different molar ratios of core histone octamer to Widom-601. Reconstitution was probed by BanI digest and analyzed by agarose gel electrophoresis.

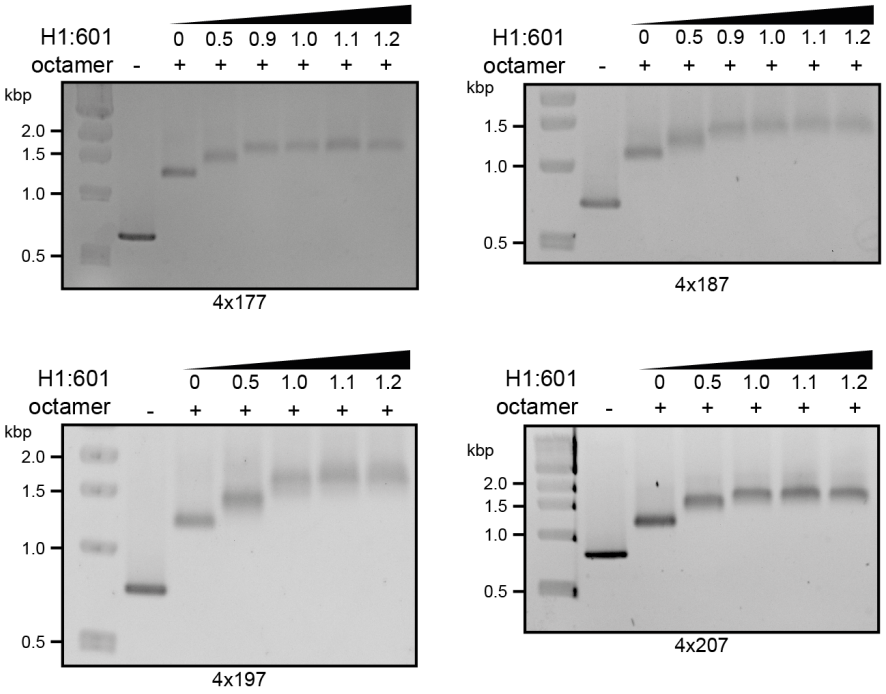

Figure 4.6: Linker histone H1 titration. Tetranucleosome arrays were reconstituted with linker histone H1 at different molar ratios of H1 to Widom-601 sequences by salt-gradient dialysis. Binding of H1 to nucleosome arrays was probed by electrophoretic mobility shift assay. Arrays saturated with H1 at circa equimolar ratios.

4.1.4 Reconstitution of tetranucleosomal arrays with H1.4

Nucleosome arrays were reconstituted by salt gradient dialysis as previously described (Dyer et al., 2004; Song et al., 2014). First, to prevent unspecific binding of histone octamer to the DNA, the proper stoichiometric ratio for the reconstitution had to be determined. For this, reconstitutions of DNA with octamer at different molar ratio of octamer to Widom-601 nucleosome positioning sequence were performed and octamer binding was probed by BanI restriction enzyme digest (Fig. 4.5).

Next, nucleosome arrays at the lowest molar ratio of octamer to nucleosome positioning sequence at which BanI digestions was not observed were reconstituted by salt gradient dialysis with different molar ratios of linker histone H1. H1 binding was probed by electrophoretic mobility shift assay (Fig. 4.6). Arrays were saturated with H1 at roughly equimolar ratios and samples at H1:Widom-601 of 1.2 were used for cryo-electron microscopy.

4.2 Single particle cryo-electron microscopy of tetranucleosome arrays

4.2.1 *Structural analysis of tetranucleosome arrays containing H1

Linker histone H1 containing tetranucleosome arrays with defined NRL were reconstituted and analyzed by single particle cryo-electron microscopy. Focused refined maps of the individual nucleosomes of the array were resolved to resolutions between 4 to 8 Å. This resolution permitted the confident interpretation of secondary structure elements of both the core histones and linker histone H1 (Figs. 4.8-4.15). All nucleosomes of the 4x177 array (Figs. 4.7-4.9) and the first three nucleosomes of the 4x187 (Figs. 4.10, 4.11), 4x197 (Figs. 4.12, 4.13) and the 4x207 (Figs. 4.14, 4.15) were resolved and showed canonical nucleosome structure (Luger et al., 1997; Tsunaka et al., 2005). The focused refined maps of individual nucleosomes were used to build atomic models of the nucleosome units based on the H1-bound mononucleosome (PDB ID 7K5Y Zhou et al. (2021a)). The overall maps of tetranucleosomes and trinucleosomes were resolved to circa 10 Å resolution. Into these intermediate resolution overall cryo-EM maps, the models of the nucleosome units were fit and the linker DNA connecting the nucleosome units was modeled (Tables 4.1-4.4).

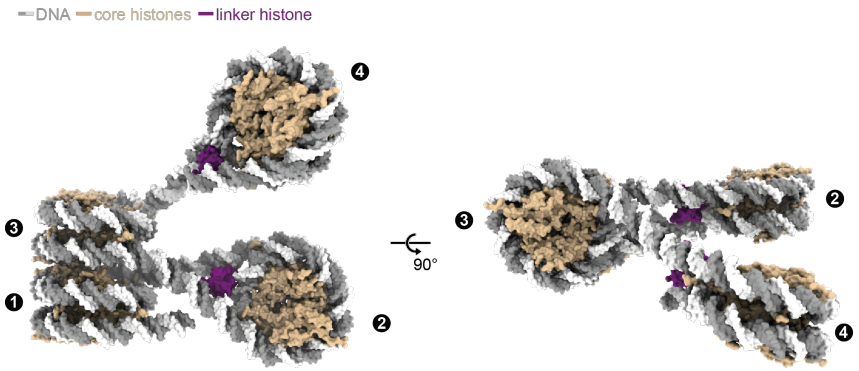

Figure 4.7: *The structure of the 4x177 tetranucleosome array containing linker histone H1. DNA containing four repeats of the Widom-601 nucleosome positioning sequence was reconstituted with human core histone octamer and human linker histone H1 and analyzed by cryo-electron microscopy. Nucleosomes of the array adopt the previously characterized zig-zag conformation (Schalch et al., 2005). Nucleosome 1 and nucleosome 3 form a stack while nucleosome 2 does not stack with nucleosome 4. All four nucleosomes and an overall map of the 4x177 could be resolved. Due to increased flexibility of the nucleosomes with increasing linker DNA length, overall tetranucleosome maps of the 4x187, 4x197 and 4x207 could not be attained. Depicted here is a surface representation of the atomic model that was built into the tetranucleosome cryo-EM density shown in Figures 4.8 and 4.9. DNA colored in grey and white, core histone octamer in wheat, linker histone H1 in purple.

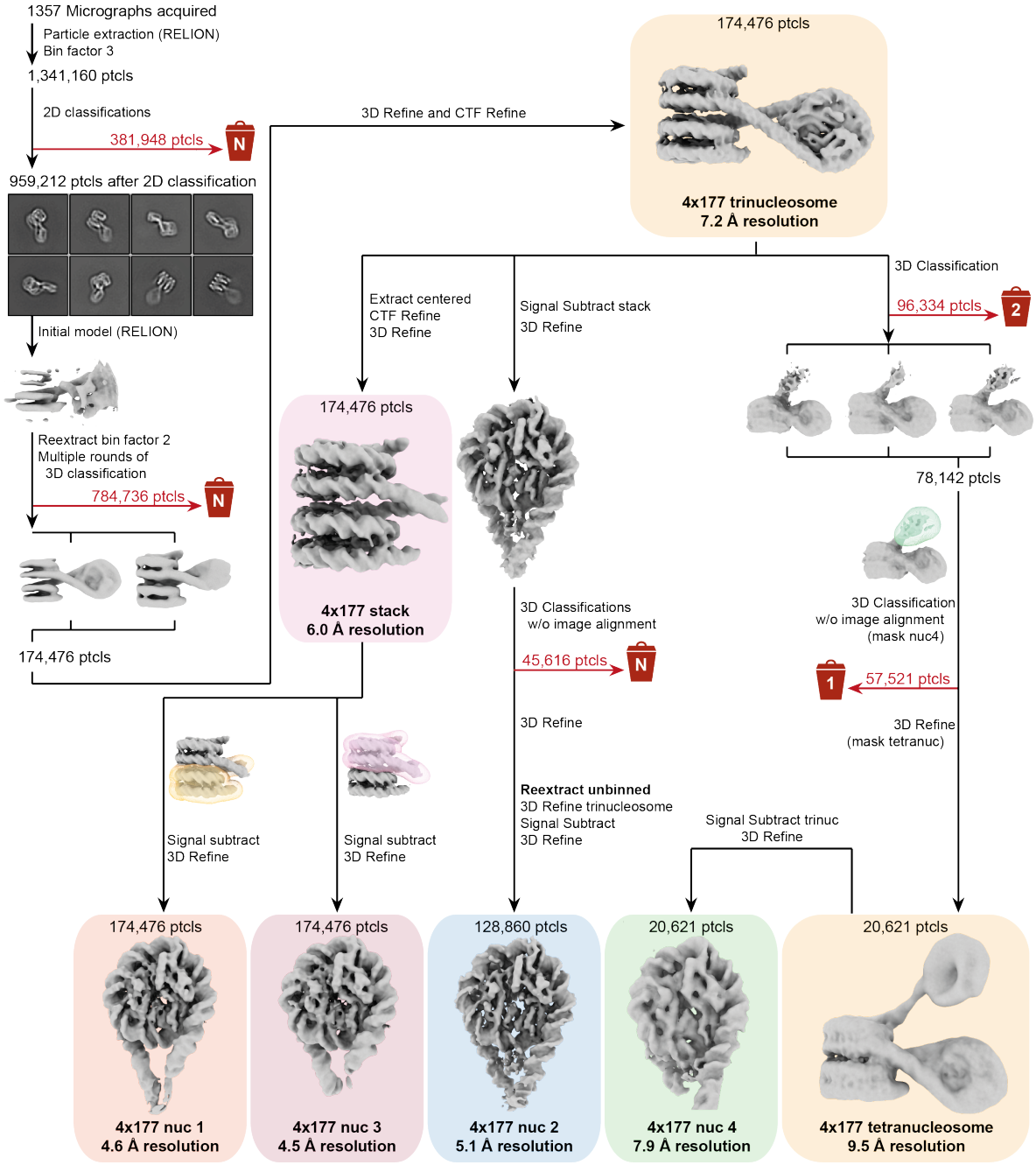

Figure 4.8: *Cryo-EM single particle analysis of 4x177. Particles were picked in Warp(Tegunov and Cramer, 2019), extracted 3x binned in RELION(Scheres, 2012) and cleaned by several rounds of 2D classification in cryoSPARC(Punjani et al., 2017). Particles belonging to classes with two or more nucleosomes visible were reextracted 2x binned and used for initial model generation in RELION. Several rounds of 3D classification selecting for trinucleosomes and 3D refinement yielded a trinucleosome map from which signal subtractions and masked classifications and refinements yielded the maps for the individual nucleosomal units and the tetranucleosome. The numbers in the bin denote the number of classes discarded during a classification and is N if more than one classification was performed.

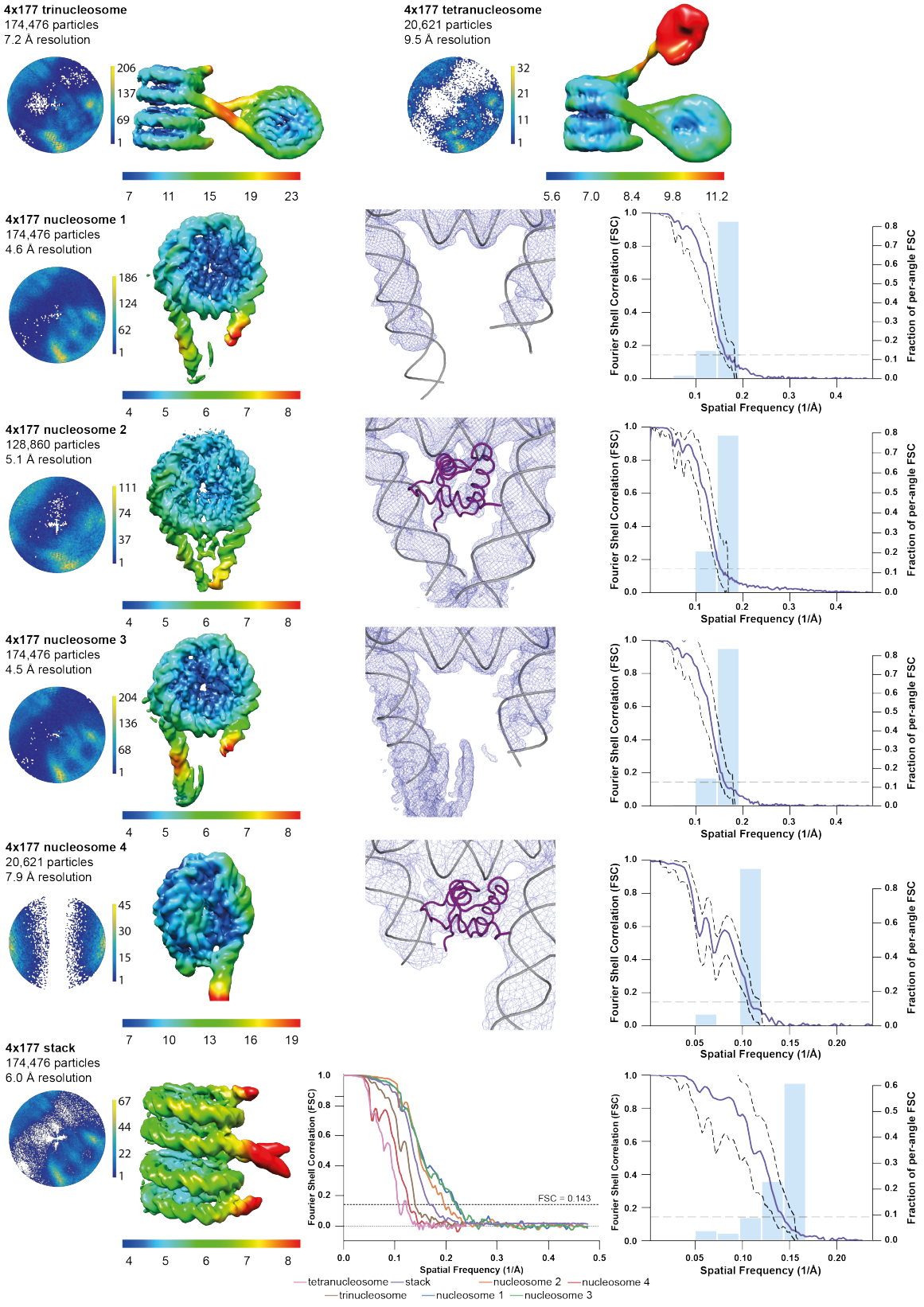

Figure 4.9: *Angular distribution, local resolution, FSCs and densities of 4x177. The cryo-EM reconstructions obtained by the image processing outlined in Figure 4.8 are colored by their local resolution determined in RELION(Scheres, 2012). Listed are the map name, number of particles used for the final reconstruction, the nominal resolution determined by RELION using the FSC = 0.143 threshold (FSC plot for all reconstructions bottom row middle panel), and a plot showing angular distribution. For the individual nucleosomal units, the middle panel shows a closeup on the nucleosome dyad with the atomic model (DNA in grey, H1 in purple), and the right panel shows the reconstruction?s 3D FSC(Tan et al., 2017).

Table 4.1: *Cryo-EM data collection, refinement and validation statistics for the 4x177 array

	tetranuc EMD-13356 PDB 7PET	**trinuc** EMD-13357 PDB 7PEU	**stack** EMD-13358 PDB 7PEV	**nuc1** EMD-13359 PDB 7PEW	**nuc2** EMD-13360 PDB 7PEX	**nuc3** EMD-13361 PDB 7PEY	**nuc4** EMD-13362 PDB 7PEZ
Data collection and processing							
Magnification	81 000x	81 000x	81 000x	81 000x	81 000x	81 000x	81 000x
Voltage (kV)	300	300	300	300	300	300	300
Electron exposure (e^{-}/Å^2)	60	60	60	60	60	60	60
Defocus range (μm)	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0
Pixel size (Å/pix)	1.05	1.05	1.05	1.05	1.05	1.05	1.05
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1
Initial particle images	1,341,160	1,341,160	1,341,160	1,341,160	1,341,160	1,341,160	1,341,160
Final particle images	20,621	174,476	174,476	174,476	128,860	174,476	20,621
Map resolution (Å) 0.143 FSC criterion	9.5	7.2	6.0	4.6	5.1	4.5	7.9
Map resolution range (Å)	6.8 - 25	5.6 - 11	4.5 - 10.8	4.2 - 8.0	4.5 - 7.3	4.1 - 8.7	7.2 - 25
Refinement							
Initial models used (PDB code)				7K5Y	7K5Y	7K5Y	7K5Y
Model resolution (Å)	9.5	7.2	6.0	4.6	5.1	4.5	7.9
Model resolution range (Å)	6.8 - 25	5.6 - 11	4.5 - 10.8	4.2 - 8.0	4.5 - 7.3	4.1 - 8.7	7.2 - 25
Map sharpening *B* factor (Å^2)	-500	-300	-310	-150	-520	-150	-300
Model composition							
Non-hydrogen atoms	54040	40119	26403	13304	13880	13099	14058
Protein residues	3222	2379	1536	768	843	768	843
DNA	1396	1040	694	352	354	342	364
B factors (Å^2)							
Protein	248	203	252	224	140	252	373
DNA	220	302	372	294	191	318	419
R.m.s. deviations							
Bond lengths (Å)	0.007	0.006	0.006	0.006	0.005	0.006	0.006
Bond angles ($^\circ$)	1.280	1.087	0.957	0.912	0.932	0.895	0.910
Validation							
MolProbity score	1.64	1.42	1.45	1.30	1.36	1.14	1.46
Clashscore	13.62	7.70	8.82	5.53	6.47	7.45	8.54
Poor rotamers (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ramachandran plot							
Favored (%)	98.4	98.6	98.4	98.9	98.4	98.4	98.2
Allowed (%)	1.6	1.4	1.6	1.1	1.6	1.6	1.8
Disallowed (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0

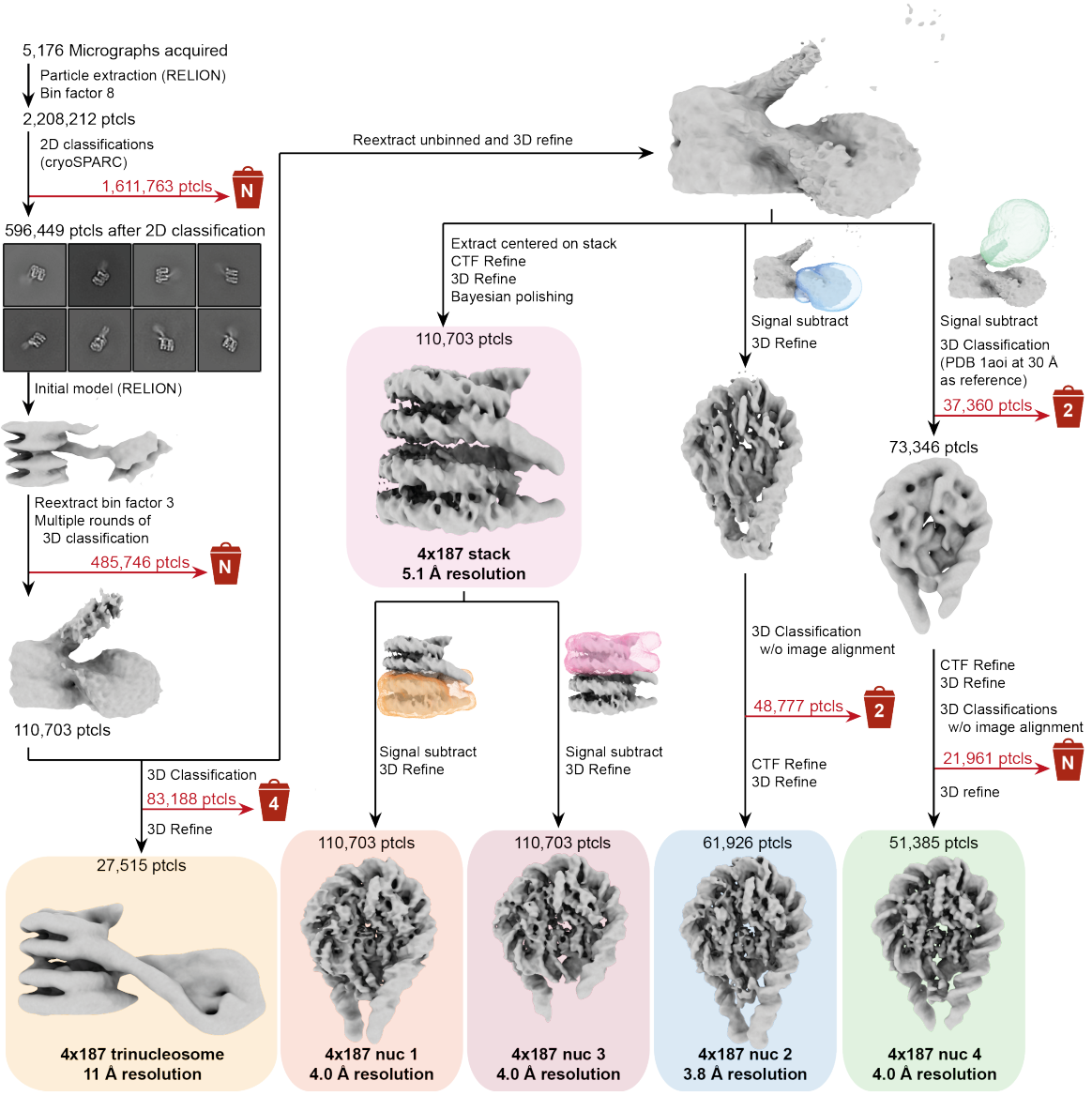

Figure 4.10: *Cryo-EM single particle analysis of 4x187. Particles were picked in Warp(Tegunov and Cramer, 2019), extracted 8x binned in RELION(Scheres, 2012) and cleaned by several rounds of 2D classification in cryoSPARC(Punjani et al., 2017). Particles belonging to classes with two or more nucleosomes visible were reextracted 3x binned and used for initial model generation in RELION. Several rounds of 3D classification selecting for trinucleosomes and 3D refinement yielded a trinucleosome map with a defined nucleosome stack and a less defined connecting nucleosome. From this particle set, (a) several rounds of 3D classification and refinement yielded the trinucleosome map and (b) particles were reextracted unbinned, refined and signal subtractions, masked classifications and refinements yielded the maps for the individual nucleosomal units. The numbers in the bin denote the number of classes discarded during a classification and is N if more than one classification was performed.

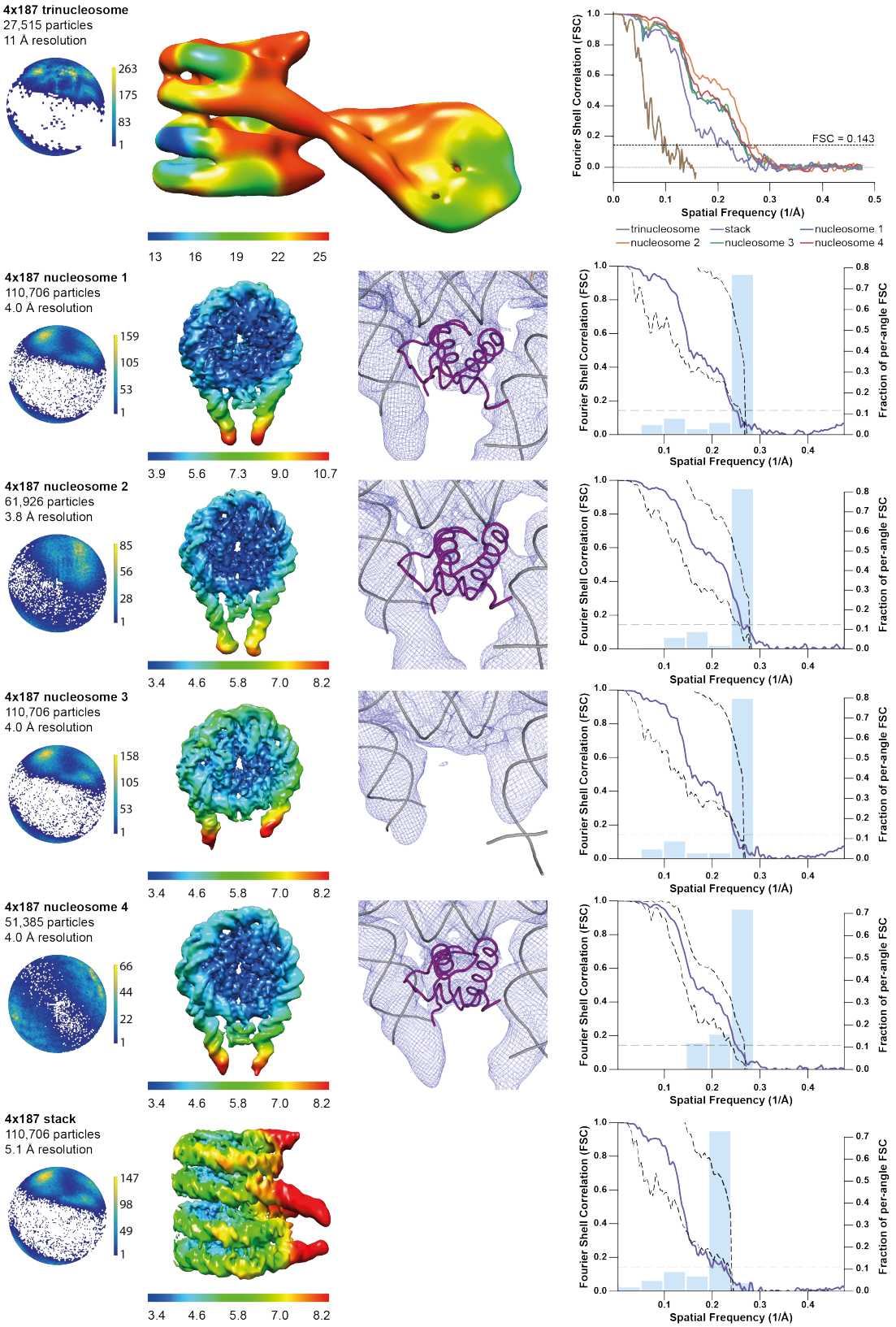

Figure 4.11: *Angular distribution, local resolution, FSCs and densities of 4x187. The cryo-EM reconstructions obtained by the image processing outlined in Figure 4.10 are colored by their local resolution determined in RELION(Scheres, 2012). Listed are the map name, number of particles used for the final reconstruction, the nominal resolution determined by RELION using the FSC = 0.143 threshold (FSC plot for all reconstructions top row right panel), and a plot showing angular distribution. For the individual nucleosomal units, the middle panel shows a closeup on the nucleosome dyad with the atomic model (DNA in grey, H1 in purple), and the right panel shows the reconstruction?s 3D FSC(Tan et al., 2017).

Table 4.2: *Cryo-EM data collection, refinement and validation statistics for the 4x187 array

	trinuc EMD-13363 PDB 7PF0	**stack** EMD-13365 PDB 7PF2	**nuc1** EMD-13369 PDB 7PF6	**nuc2** EMD-13368 PDB 7PF5	**nuc3** EMD-13367 PDB 7PF4	**nuc4** EMD-13366 PDB 7PF3
Data collection and processing						
Magnification	81 000x	81 000x	81 000x	81 000x	81 000x	81 000x
Voltage (kV)	300	300	300	300	300	300
Electron exposure ($e^-/Å^2$)	60	60	60	60	60	60
Defocus range (μm)	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0
Pixel size (Å/pix)	1.05	1.05	1.05	1.05	1.05	1.05
Symmetry imposed	C1	C1	C1	C1	C1	C1
Initial particle images	1,259,654	1,259,654	1,259,654	1,259,654	1,259,654	1,259,654
Final particle images	27,515	110,706	110,706	61,926	110,706	51,385
Map resolution (Å) 0.143 FSC criterion	11	5.1	4.0	3.8	4.0	4.0
Map resolution range (Å)		4.0 - 13	3.9 - 11.3	3.4 - 8.1	3.8 - 9.2	3.6 - 9.3
Refinement						
Initial models used (PDB code)			7K5Y	7K5Y	7K5Y	7K5Y
Model resolution (Å)	11	5.1	4.0	3.8	4.0	4.0
Model resolution range (Å)		4.0 - 13	3.9 - 11.3	3.4 - 8.1	3.8 - 9.2	3.6 - 9.3
Map sharpening *B* factor ($Å^2$)	0	-50	0	-50	0	-50
Model composition						
Non-hydrogen atoms	41515	36610	13470	13470	12935	13470
Protein residues	2454	1611	843	843	768	843
DNA	1082	678	334	334	334	334
B factors ($Å^2$)						
Protein	163	217	180	152	158	178
DNA	340	325	222	190	241	220
R.m.s. deviations						
Bond lengths (Å)	0.011	0.007	0.008	0.006	0.005	0.004
Bond angles (°)	1.335	0.974	1.054	0.889	0.887	0.944
Validation						
MolProbity score	1.56	1.47	1.62	1.32	1.41	1.43
Clashscore	9.13	8.76	9.20	5.86	5.99	6.15
Poor rotamers (%)	0.0	0.0	0.0	0.0	0.0	0.0
Ramachandran plot						
Favored (%)	97.7	98.2	97.3	98.2	97.6	97.6
Allowed (%)	2.3	1.8	2.7	1.8	2.4	2.4
Disallowed (%)	0.0	0.0	0.0	0.0	0.0	0.0

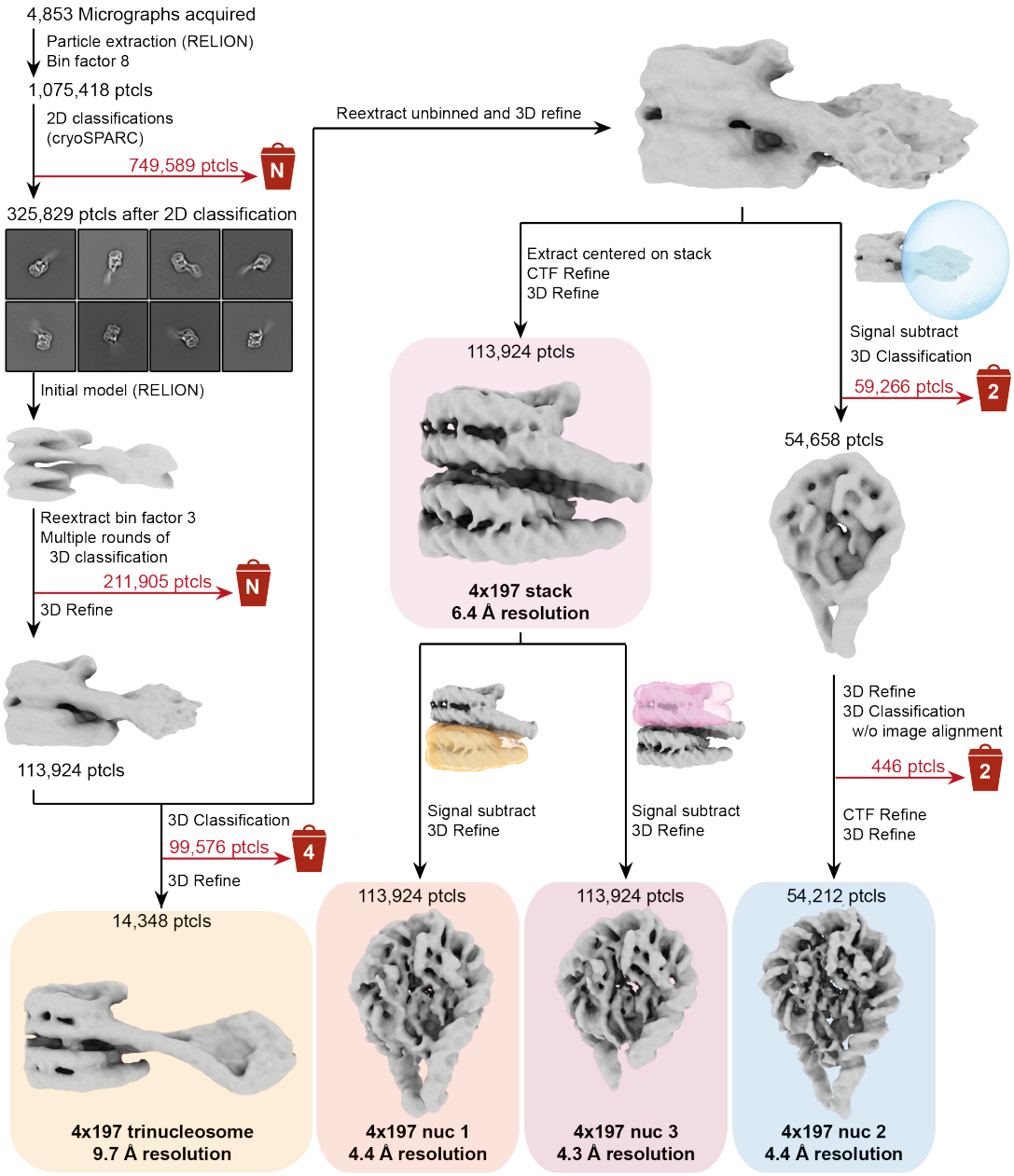

Figure 4.12: *Cryo-EM single particle analysis of 4x197. Particles were picked in Warp(Tegunov and Cramer, 2019), extracted 8x binned in RELION(Scheres, 2012) and cleaned by several rounds of 2D classification in cryoSPARC(Punjani et al., 2017). Particles belonging to classes with two or more nucleosomes visible were reextracted 3x binned and used for initial model generation in RELION. Several rounds of 3D classification selecting for trinucleosomes and 3D refinement yielded a trinucleosome map with a defined nucleosome stack and a less defined connecting nucleosome. From this particle set, (a) several rounds of 3D classification and refinement yielded the trinucleosome map and (b) particles were reextracted unbinned, refined and signal subtractions, masked classifications and refinements yielded the maps for the individual nucleosomal units. The numbers in the bin denote the number of classes discarded during a classification and is N if more than one classification was performed.

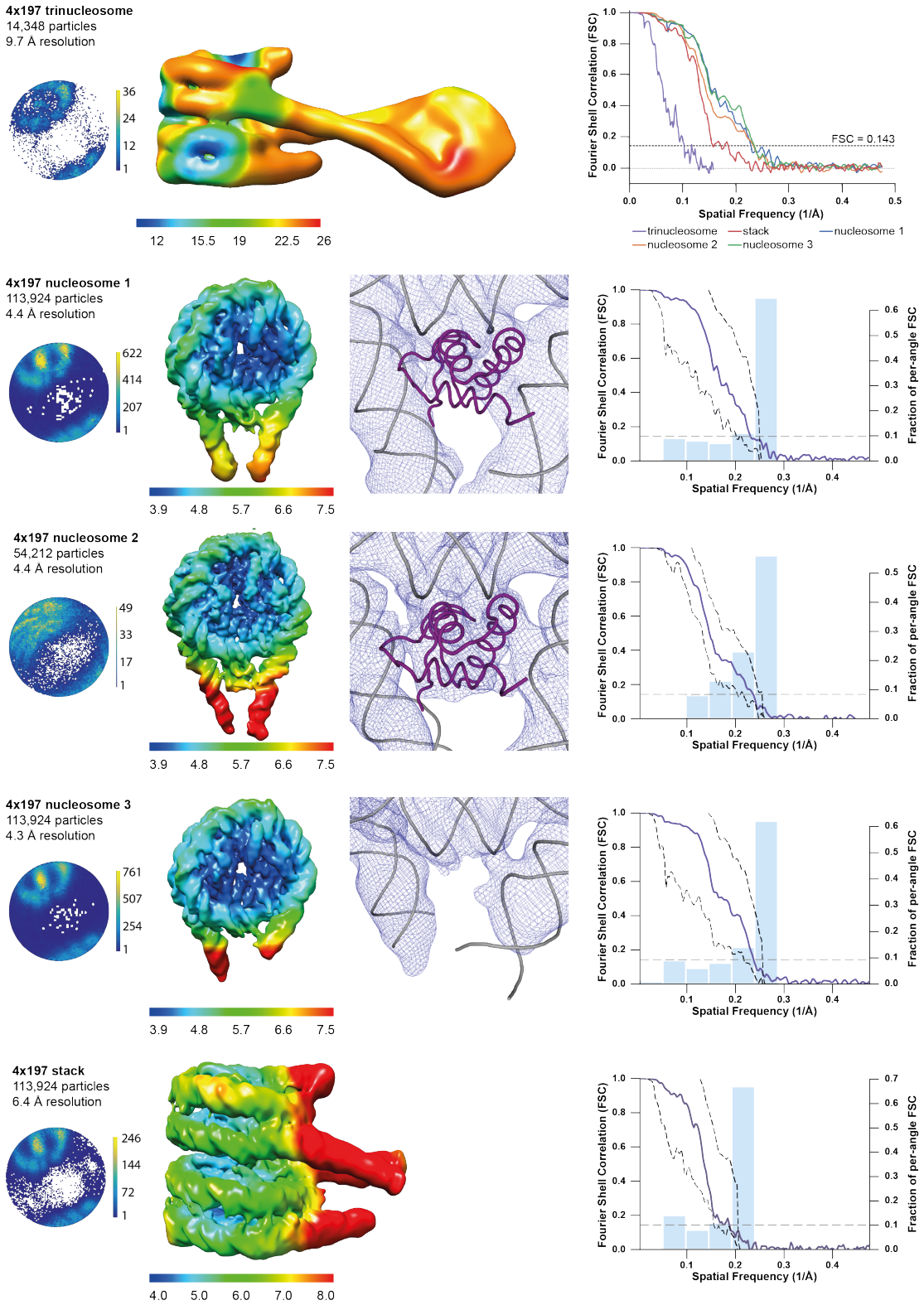

Figure 4.13: *Angular distribution, local resolution, FSCs and densities of 4x197. The cryo-EM reconstructions obtained by the image processing outlined in Figure 4.12 are colored by their local resolution determined in RELION(Scheres, 2012). Listed are the map name, number of particles used for the final reconstruction, the nominal resolution determined by RELION using the FSC = 0.143 threshold (FSC plot for all reconstructions top row right panel), and a plot showing angular distribution. For the individual nucleosomal units, the middle panel shows a closeup on the nucleosome dyad with the atomic model (DNA in grey, H1 in purple), and the right panel shows the reconstruction?s 3D FSC(Tan et al., 2017).

Table 4.3: *Cryo-EM data collection, refinement and validation statistics for the 4x197 array

	trinuc EMD-13370 PDB 7PFA	**stack** EMD-13371 PDB 7PFC	**nuc1** EMD-13372 PDB 7PFD	**nuc2** EMD-13373 PDB 7PFE	**nuc3** EMD-13374 PDB 7PFF
Data collection and processing					
Magnification	81 000x	81 000x	81 000x	81 000x	81 000x
Voltage (kV)	300	300	300	300	300
Electron exposure ($e^-/Å^2$)	60	60	60	60	60
Defocus range (μm)	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0
Pixel size (Å/pix)	1.05	1.05	1.05	1.05	1.05
Symmetry imposed	C1	C1	C1	C1	C1
Initial particle images	1,075,418	1,075,418	1,075,418	1,075,418	1,075,418
Final particle images	14,348	113,924	113,924	54,212	113,924
Map resolution (Å) 0.143 FSC criterion	9.7	6.4	4.4	4.4	4.3
Map resolution range (Å)		4.4 - 12	4.0 - 7.5	3.9 - 10	3.9 - 8.5
Refinement					
Initial models used (PDB code)			7K5Y	7K5Y	7K5Y
Model resolution (Å)	9.7	6.4	4.4	4.4	4.3
Model resolution range (Å)		4.4 - 12	4.0 - 7.5	3.9 - 10	3.9 - 8.5
Map sharpening *B* factor ($Å^2$)	0	-50	-50	-100	-50
Model composition					
Non-hydrogen atoms	42335	26814	13675	13880	12935
Protein residues	2454	1611	843	843	768
DNA	1122	688	334	354	334
B factors ($Å^2$)					
Protein	324	516	408	202	366
DNA	391	589	421	252	413
R.m.s. deviations					
Bond lengths (Å)	0.007	0.006	0.006	0.007	0.006
Bond angles (°)	1.137	0.950	0.915	0.936	0.921
Validation					
MolProbity score	1.50	1.59	1.44	1.39	1.49
Clashscore	9.43	10.58	7.78	7.08	9.19
Poor rotamers (%)	0.0	0.0	0.0	0.0	0.0
Ramachandran plot					
Favored (%)	98.3	97.8	98.0	98.5	98.4
Allowed (%)	1.7	2.2	2.0	1.5	1.6
Disallowed (%)	0.0	0.0	0.0	0.0	0.0

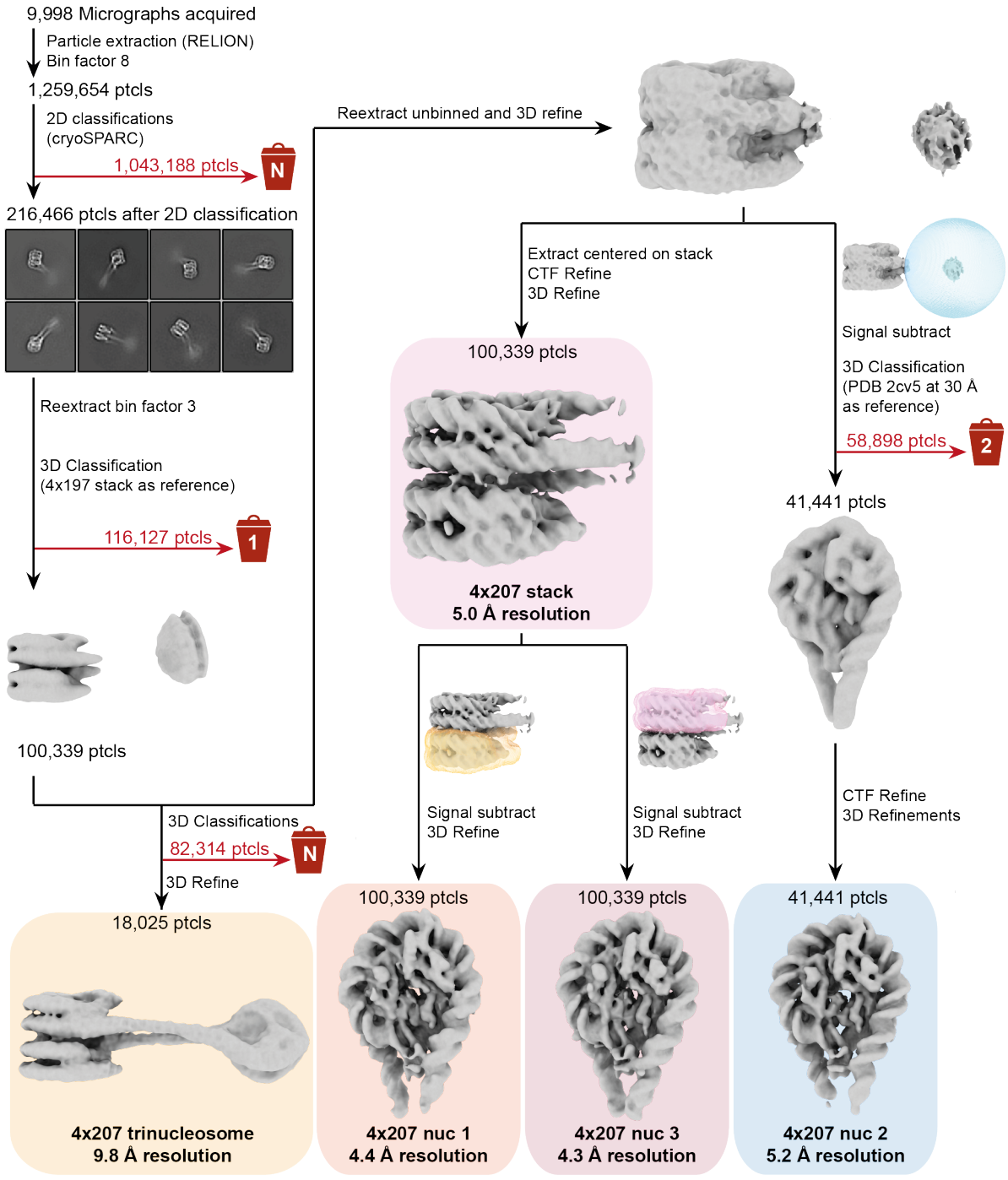

Figure 4.14: *Cryo-EM single particle analysis of 4x207. Particles were picked in Warp(Tegunov and Cramer, 2019), extracted 8x binned in RELION(Scheres, 2012) and cleaned by several rounds of 2D classification in cryoSPARC(Punjani et al., 2017). Particles belonging to classes with two or more nucleosomes visible were reextracted 3x binned. Several rounds of 3D classification against the lowpass filtered 4x197 stack map selecting for nucleosome stacks and 3D refinement yielded a trinucleosome map with a defined nucleosome stack and an ill-defined connecting nucleosome. From this particle set, (a) several rounds of 3D classification and refinement yielded the trinucleosome map and (b) particles were reextracted unbinned, refined and signal subtractions, masked classifications and refinements yielded the maps for the individual nucleosomal units. The numbers in the bin denote the number of classes discarded during a classification and is N if more than one classification was performed.

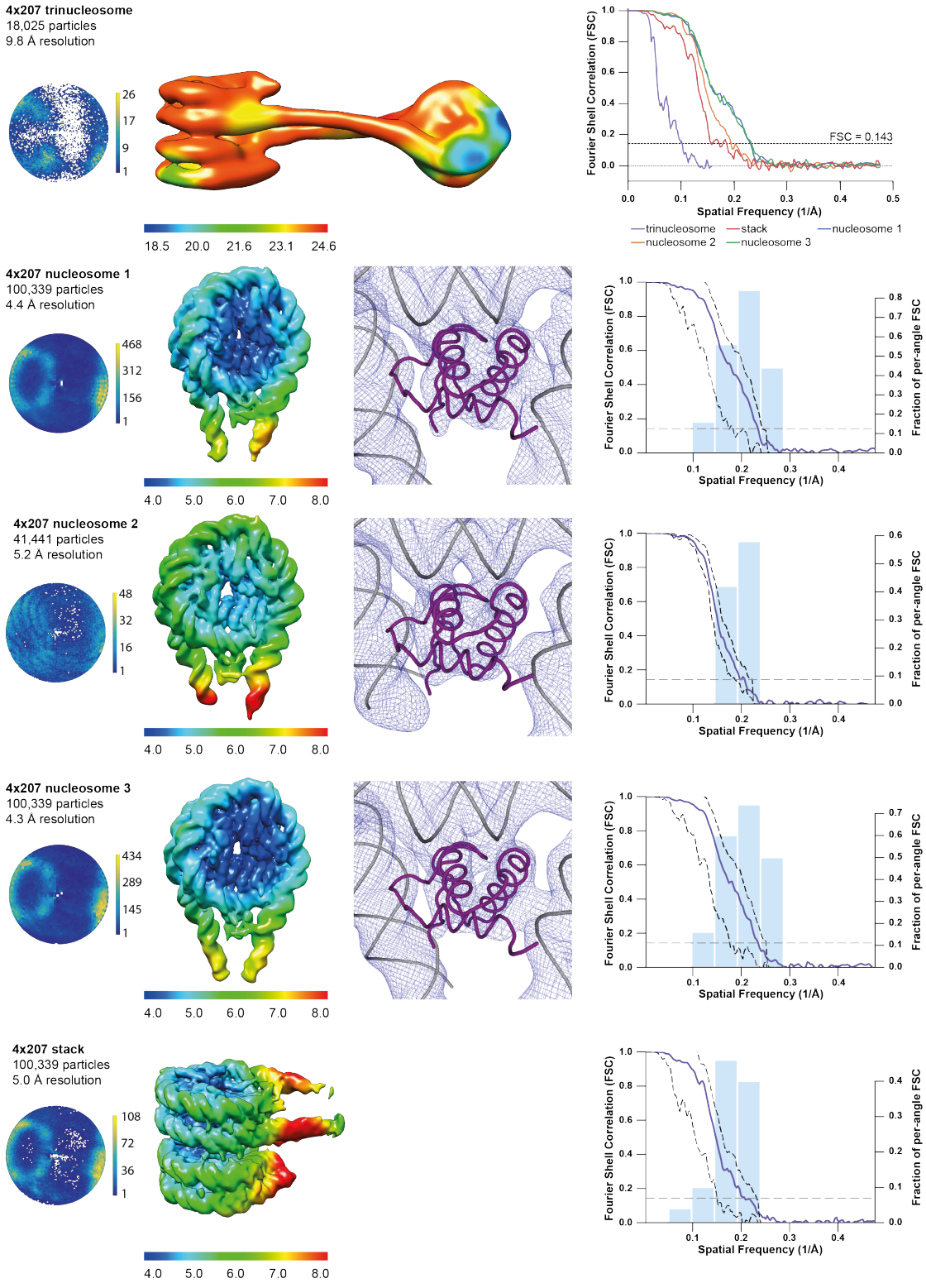

Figure 4.15: *Angular distribution, local resolution, FSCs and densities of 4x207. The cryo-EM reconstructions obtained by the image processing outlined in Figure 4.14 are colored by their local resolution determined in RELION(Scheres, 2012). Listed are the map name, number of particles used for the final reconstruction, the nominal resolution determined by RELION using the FSC = 0.143 threshold (FSC plot for all reconstructions top row right panel), and a plot showing angular distribution. For the individual nucleosomal units, the middle panel shows a closeup on the nucleosome dyad with the atomic model (DNA in grey, H1 in purple), and the right panel shows the reconstruction?s 3D FSC(Tan et al., 2017).

Table 4.4: *Cryo-EM data collection, refinement and validation statistics for the 4x207 array

	trinuc EMD-13379 PDB 7PFT	**stack** EMD-13380 PDB 7PFU	**nuc1** EMD-13381 PDB 7PFV	**nuc2** EMD-13382 PDB 7PFW	**nuc3** EMD-13383 PDB 7PFX
Data collection and processing					
Magnification	81 000x	81 000x	81 000x	81 000x	81 000x
Voltage (kV)	300	300	300	300	300
Electron exposure ($e^-/Å^2$)	60	60	60	60	60
Defocus range (μm)	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0	0.5-2.0
Pixel size (Å/pix)	1.05	1.05	1.05	1.05	1.05
Symmetry imposed	C1	C1	C1	C1	C1
Initial particle images	1,259,654	1,259,654	1,259,654	1,259,654	1,259,654
Final particle images	18,025	100,339	100,339	41,441	100,339
Map resolution (Å) 0.143 FSC criterion		5.0	4.4	5.2	4.3
Map resolution range (Å)		4.1 - 7.6	4.2 - 7.4	4.5 - 9.9	4.1 - 7.6
Refinement					
Initial models used (PDB code)			7K5Y	7K5Y	7K5Y
Model resolution (Å)		5.0	4.4	5.2	4.3
Model resolution range (Å)	4.1 - 7.6	4.2 - 7.4	4.5 - 9.9	4.1 - 7.6	
Map sharpening *B* factor ($Å^2$)	0	-100	-100	-100	-100
Model composition					
Non-hydrogen atoms	44100	27760	13880	13470	13880
Protein residues	2529	1686	843	843	843
DNA	1182	708	351	334	354
B factors ($Å^2$)					
Protein	344	403	325	376	330
DNA	549	474	384	382	380
R.m.s. deviations					
Bond lengths (Å)	0.006	0.006	0.006	0.006	0.006
Bond angles (°)	1.026	0.894	0.864	0.920	0.880
Validation					
MolProbity score	1.42	1.47	1.44	1.43	1.39
Clashscore	7.69	8.78	8.13	7.86	7.13
Poor rotamers (%)	0.0	0.0	0.0	0.0	0.0
Ramachandran plot					
Favored (%)	98.7	98.1	98.8	98.9	98.7
Allowed (%)	1.3	1.9	1.2	1.1	1.3
Disallowed (%)	0.0	0.0	0.0	0.0	0.0

4.2.2 *Overall structure of tetranucleosome arrays

The four studied H1 containing tetranucleosome arrays adopt a zig-zag arrangement of nucleosomes (Figures 4.7, 4.16) which was previously described for the 4x167 crystal structure without H1 (Schalch et al., 2005), the 6x187 crystal structure with H1 (Garcia-Saez et al., 2018), and the cryo-EM structures of the 12x187 and 12x197 containing H1 (Song et al., 2014). Nucleosomes 1 and 3 form a canonical stack (Schalch et al., 2005) with nucleosome 2 looped out and rotated by circa 70-90° relative to the nucleosome stack (Figs. 4.7, 4.16). The distance between nucleosome 2 increases with increasing NRL (Fig. 4.16). The increasing length of linker DNA leads to higher flexibility of nucleosome 2 with respect to the nucleosome stack, as can be seen in the fuzzy nature of this nucleosome in the 2D classes and during 3D processing (Figs 4.8, 4.10, 4.12, 4.14).

Nucleosome 4 does not stack with nucleosome 2 (Fig. 4.7) and is thus flexible as well with increased mobility as the DNA linker gets longer. Thus, the 4x177 tetranucleosome is only resolved at intermediate resolution (Figs. 4.8, 4.9) and the 4x187 shows only very weak density for nucleosome 4 in the full tetranucleosome map (Fig. 4.10). Nevertheless, nucleosome 4 of the 4x187 array could be resolved separately by focused refinement (Figs. 4.10, 4.11). However, nucleosome 4 was too mobile to be resolved either in the overall map or separately in the 4x197 (Figs. 4.12, 4.13) and 4x207 (Figs. 4.14, 4.15) arrays. Nonetheless, the overall architecture of the tetranucleosome arrays appears to be very similar, as the trajectory of the linker DNA connecting nucleosome 3 and nucleosome 4 is the same in the 4x177 (Fig. 4.8) and 4x187 (Fig. 4.10), indicating that it is likely to be similar also in the 4x197 and 4x207 arrays as the arrangement of nucleosome stack with looped out and non-stacking nucleosome 2 is shared between all arrays (Fig. 4.16).

4.2.3 *Nucleosome stacking in solution

Two types of nucleosome stacking in nucleosome arrays have been characterized at high resolution by previous structural studies (Song et al., 2014; Garcia-Saez et al., 2018). In type I interactions, nucleosomes closely pack with contacts between H2A-H2B dimers of apposing nucleosomes. These interactions have been observed in the crystal structure of 4x167 without H1 (Schalch et al., 2005) and within the tetranucleosome units of the cryo-EM structures of the 12x177 and 12x187 array containing H1 (Song et al., 2014) (Fig. 4.17). Nucleosome are slightly offset and stack more loosely in type II interactions. Here, the N-terminal tail of H4 interacts with the acidic patch of the apposing nucleosome, as observed for the crystal structure of the 6x187 with H1 (Garcia-Saez et al., 2018) and for the contacts between tetranucleosome units of the cryo-EM structures of the 12x177 and 12x187 with H1 (Song et al., 2014) (Fig. 4.17). Additionally, other interactions between nucleosomes have been observed for mononucleosomes in crystals or by cryo-EM in solution (Bilokapic et al., 2018).

Nucleosomes 1 and 3 in the arrays studied here form a compact stack with contacting H2A-H2B dimers that closely resembles the type I interactions (Fig. 4.17). Whereas internucleosome interactions appear to be similar to those of type I interactions, the dyad axes of nucleosomes 1 and 3 are aligned almost parallel compared to type I interactions where dyad axes are slightly tilted toward each other (Fig. 4.17). This might be due to absence of H1 in the 4x167 crystal structure (Schalch et al., 2005) or the difference of H1 binding mode in the 12x177 and 12x187 cryo-EM structures (Song et al., 2014).

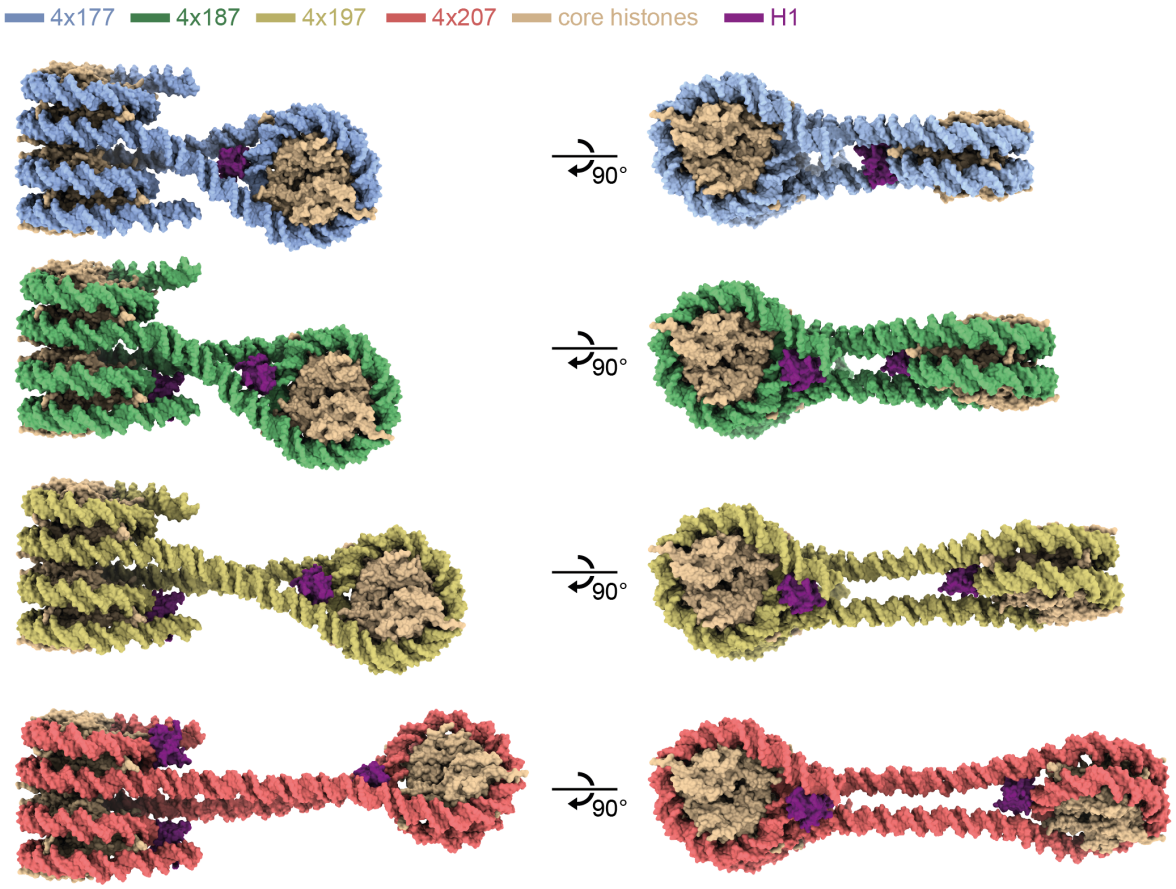

Figure 4.16: *The structure of trinucleosome cores of tetranucleosome arrays. The trinucleosome cores of the 4x177, 4x187, 4x197 and 4x207 show nucleosome 2 rotated relative to the nucleosome stack. With increasing length of linker DNA, nucleosome 2 moves further away from the nucleosome stack. NRL 177 in blue, NRL 187 in green, NRL 197 in yellow, NRL 207 in red, core histones in wheat, linker histone H1 in purple.

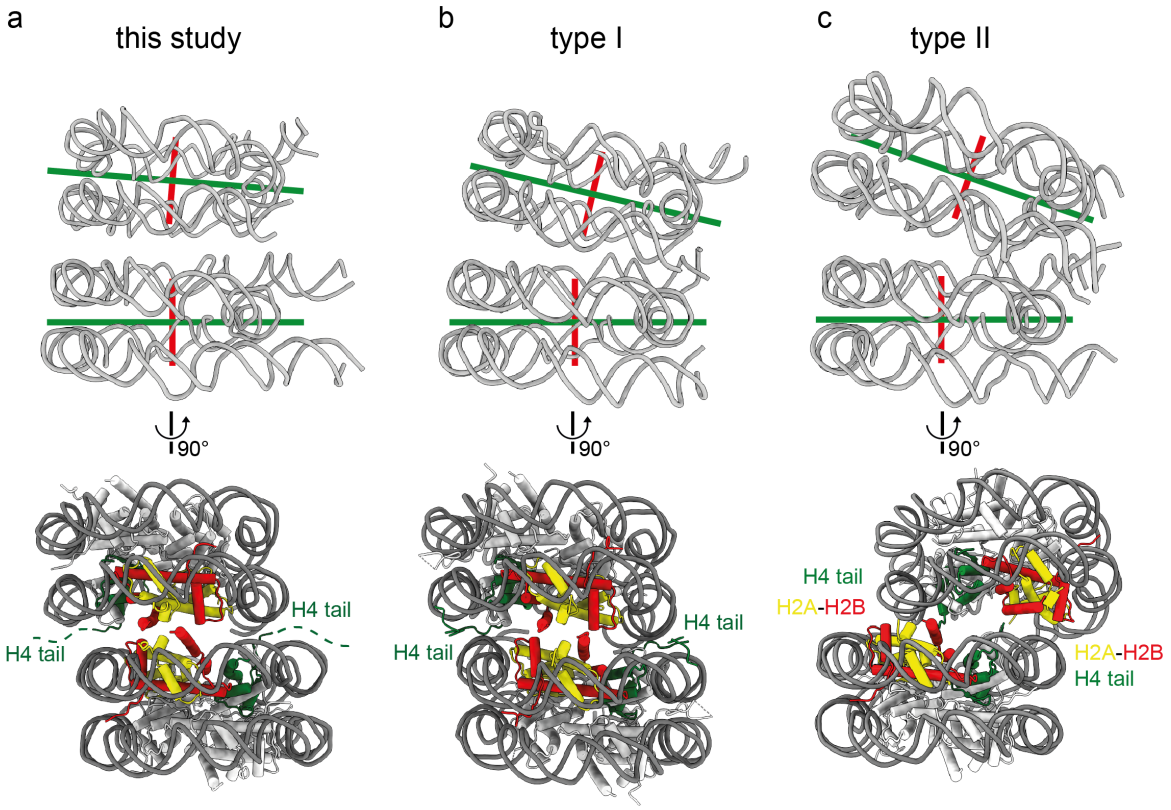

Figure 4.17: *Nucleosome stacking in nucleosome arrays. Nucleosome stacking in tetranucleosome arrays is similar to the stacking observed in the crystal structure of the 4x167 array without H1 (Schalch et al., 2005) and the cryo-EM reconstruction of the 12x177 and 12x187 array with H1 (Song et al., 2014). **(a)** Nucleosome stack from the 4x187 array represents stacks from all structures reported in this study. **(b)** nucleosome stack from the 4x167 crystal structure (PDB ID 1ZBB (Schalch et al., 2005)) represents the type I interaction observed in the 4x167 crystal structure and within tetranucleosomal units of the 12x177 and 12x187 cryo-EM structure (Song et al., 2014). **(c)** nucleosome stack from the 6x187 crystal structure (PDB ID 6HKT (Garcia-Saez et al., 2018)) represents the type II interaction observed also between tetranucleosome units of the 12x177 and 12x187 cryo-EM structure. Top: Dyad axes drawn in green run almost parallel in the stacking observed in the cryo-EM reconstructions determined for 4x177, 4x187, 4x197 and 4x207 whereas dyad axes in the stack observed in both type I and type II interactions are slightly tilted toward each other. Bottom: The interface between stacking nucleosomes in the 4x177, 4x187, 4x197, 4x207 reported here and in type I interactions consists of apposed H2A-H2B dimers (H2A in yellow, H2B in red) while in type II interactions the nucleosome stack is slightly offset and places the N-terminal part of H4 (green) near the H2A-H2B dimer.

4.2.4 *H1 orientation and DNA interactions

Linker histone H1 is bound on the nucleosome dyad (Figs. 4.18, 4.9, 4.11, 4.13, 4.15). In the ten focused-refined maps where linker histone H1 is bound, H1 contacts nucleosomal and linker DNA with the three previously characterized contacts (Zhou et al., 2015; Bednar et al., 2017; Zhou et al., 2021a) (Fig. 4.18a). Loop L3 and the N-terminal part of helix α2 contact nucleosomal DNA at the dyad, while helix α3 binds one linker DNA and loop L1 binds the other linker DNA (Figs. 4.18a, 4.9, 4.11, 4.13, 4.15). This mode is described in the literature as "on-dyad" (Zhou et al., 2015, 2021a) or, more accurately, as "lopsided" (Bednar et al., 2017) because H1 is shifted slightly off-dyad. In the arrays H1 is arranged on nucleosome 1, 2 and 4 to bind entering linker DNA with its α3 helix while on nucleosome 3 the α3 helix is bound to the DNA exiting the nucleosome (Fig. 4.18b). H1 thus appears to be orientable to contact entry or exit DNA depending on local DNA geometry. The orientation of H1 influences which part of the protein is accessible for post-translational modifications or binding by other factors. The orientation on nucleosome stacks exposes the unstructured N-terminal tail containing residues such as K26, S27, K34, and S35 which have been implicated in regulating H1 mobility (Kamieniarz et al., 2012; Chu et al., 2011) and heterochromatin formation (Daujat et al., 2005) (Fig. 4.19).

4.2.5 *H1 binding depends on nucleosome repeat length

While the overall architecture of the tetranucleosome arrays studied here appears to be conserved, the structures differ in linker histone H1 binding to the individual nucleosomes of the array (Figure 4.18b). H1 is bound to nucleosome 2 in all four structures and is observed on nucleosome 4 in those arrays where nucleosome 4 is resolved. However, H1 presence on stacked nucleosomes differs between the four arrays. H1 is absent from both stacking nucleosomes of the 4x177, while it is present on nucleosome 1 in the 4x187 and 4x197 and on both stacked nucleosomes in the 4x207. Thus, all non-stacking nucleosomes of the arrays are bound by H1 but H1 binding to stacked nucleosomes occurs only with increasing NRL (Fig. 4.18b).

4.2.6 *H1 binding depends on linker DNA trajectory

As H1 binds to the nucleosomes of the array in the canonical manner with three DNA contacts and as H1 is absent from some stacking nucleosomes, it seemed likely that DNA

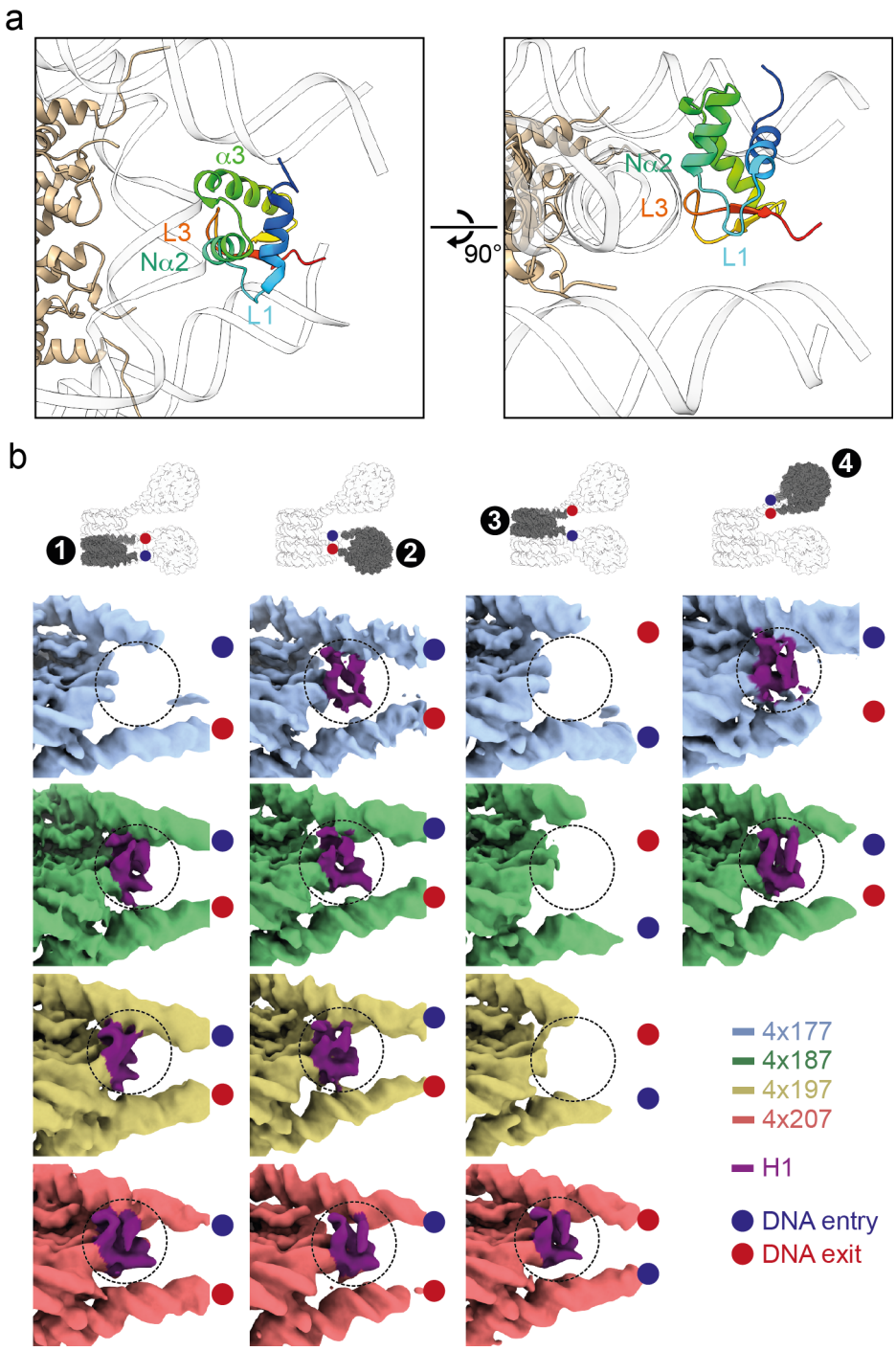

Figure 4.18: *H1 binding depends on nucleosome repeat length. (a) H1 binds to nucleosomes of the array near the nucleosome dyad. The N-terminal part of the α2-helix (Nα2) and the L3 loop contact the DNA around the dyad, whereas the α3-helix and the L1 loop interact with linker DNAs. H1 rainbow-colored from N to C terminus, DNA in white, histone octamer in wheat. **(b)** Focused-refined cryo-EM densities for nucleosomes 1, 2, 3 and 4 colored by NRL (4x177 blue, 4x187 green, 4x197 yellow, 4x207 red). H1 density is in purple. Nucleosomes are all viewed the same way. Entry and exit DNA are marked by a blue and a red dot, respectively. Focused refined maps of nucleosome 4 could not be obtained for the 4x197 and 4x207 arrays due to higher mobility.

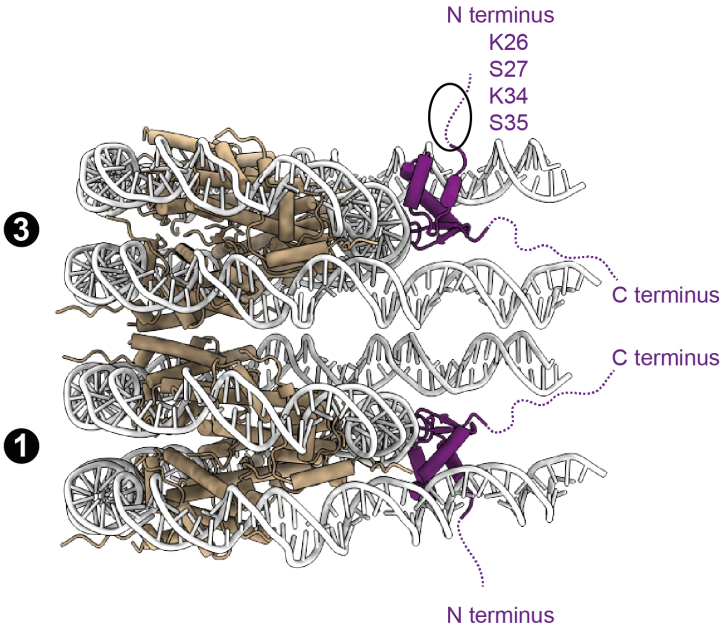

Figure 4.19: *H1 N-terminal regions extend from nucleosome stack in opposite directions. Residues regulating H1 mobility (K34 (Kamieniarz et al., 2012) and S35 (Chu et al., 2011)) and heterochromatin formation (K26 and S27)(Daujat et al., 2005) protrude from the nucleosome stack on both sides and are accessible for protein-protein interactions. The first ordered N-terminal residue of H1 is S35, disordered residues were not observed and indicated by a dashed line. DNA in grey, histone octamer in wheat, H1 in purple.

geometry influences whether H1 can bind to the nucleosomes by removing interaction surfaces. An overlay of the structure of the four arrays revealed a stark and progressive change in linker DNA trajectory with increasing NRL (Fig. 4.20). Linker DNA geometry can be described by the angle α between the nucleosome dyad axis and linker DNA measured in the plane of the nucleosome disc and by the angle β between the nucleosome dyad axis and linker DNA measured in the plane normal to the nucleosome disc along the dyad axis (Bednar et al., 2017). Due to 90° rotation between nucleosome 2 and the nucleosome stack (Fig. 4.16), β seems more likely to influence H1 binding. To describe α and β relative to the isolated H1 bound mononucleosome (PDB ID 7K5Y), $\Delta\alpha$ and $\Delta\beta$ as the difference in α and β, respectively, between the nucleosomes of the array and the H1 bound mononucleosome were calculated (Table 4.5).

As anticipated, $\Delta\beta$ appeared to correlate with linker histone H1 binding (Fig. 4.21) and nucleosomes with a $\Delta\beta$ close to zero for both linker DNAs had H1 bound (Fig. 4.21a). Low $\Delta\beta$ values were observed for nucleosomes 2 and 4 of all arrays where linker histone H1 was bound in all arrays (Table 4.5). However, nucleosomes with high $\Delta\beta$ values did not have H1 bound, likely because sufficient contacts between linker histone and linker

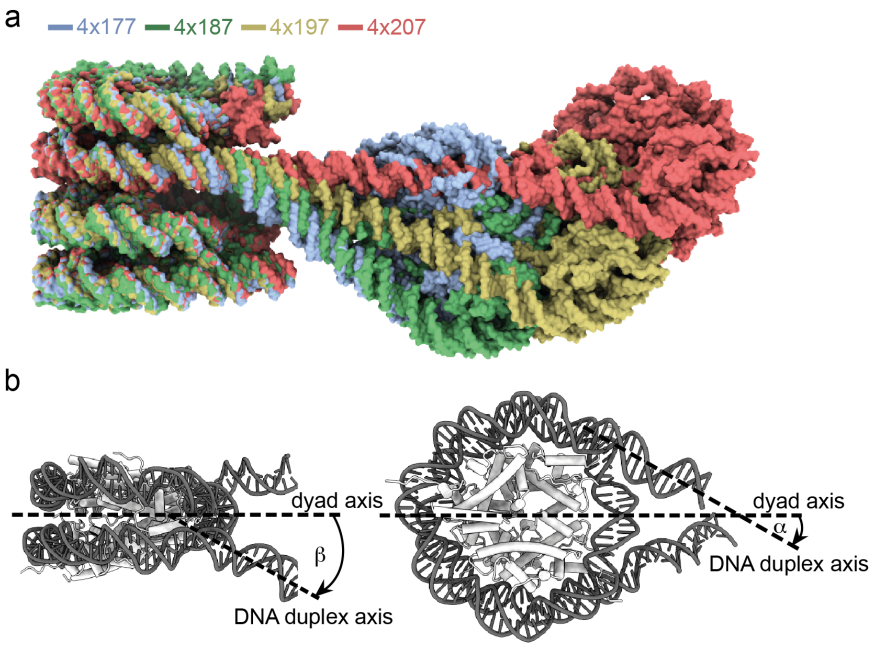

Figure 4.20: *NRL alters linker DNA trajectory at stacked nucleosomes. (a) Overlay of all four trinucleosome structures shown in Fig. reftrinucs. With increasing NRL, linker DNA trajectories at the stacked nucleosomes are altered. **(b)** β is defined as the angle between the nucleosome dyad and the linker DNA duplex axis, projected onto the plane perpendicular to the nucleosome disc (Bednar et al., 2017). α is defined as the angle between the nucleosome dyad and the linker DNA duplex axis, projected onto the plane of to the nucleosome disc.

DNA could not be formed. High $\Delta\beta$ values were observed for entry DNA at nucleosome 3, except for the 4x207 that is also the only array with H1 bound to this nucleosome (Fig. 4.21b). Similarly, for nucleosome 1 the highest $\Delta\beta$ was observed in the 4x177 array that is also the only array with H1 absent on this nucleosome (Fig. 4.21c). Thus, as nucleosome 2 moves further away from the nucleosome stack, linker DNA trajectories at nucleosomes 1 and 3 progressively relax and approach that of the isolated H1 bound mononucleosome (Fig. 4.21a). This enables contacts between H1 and linker DNA in arrays with long NRLs (Fig. 4.22).

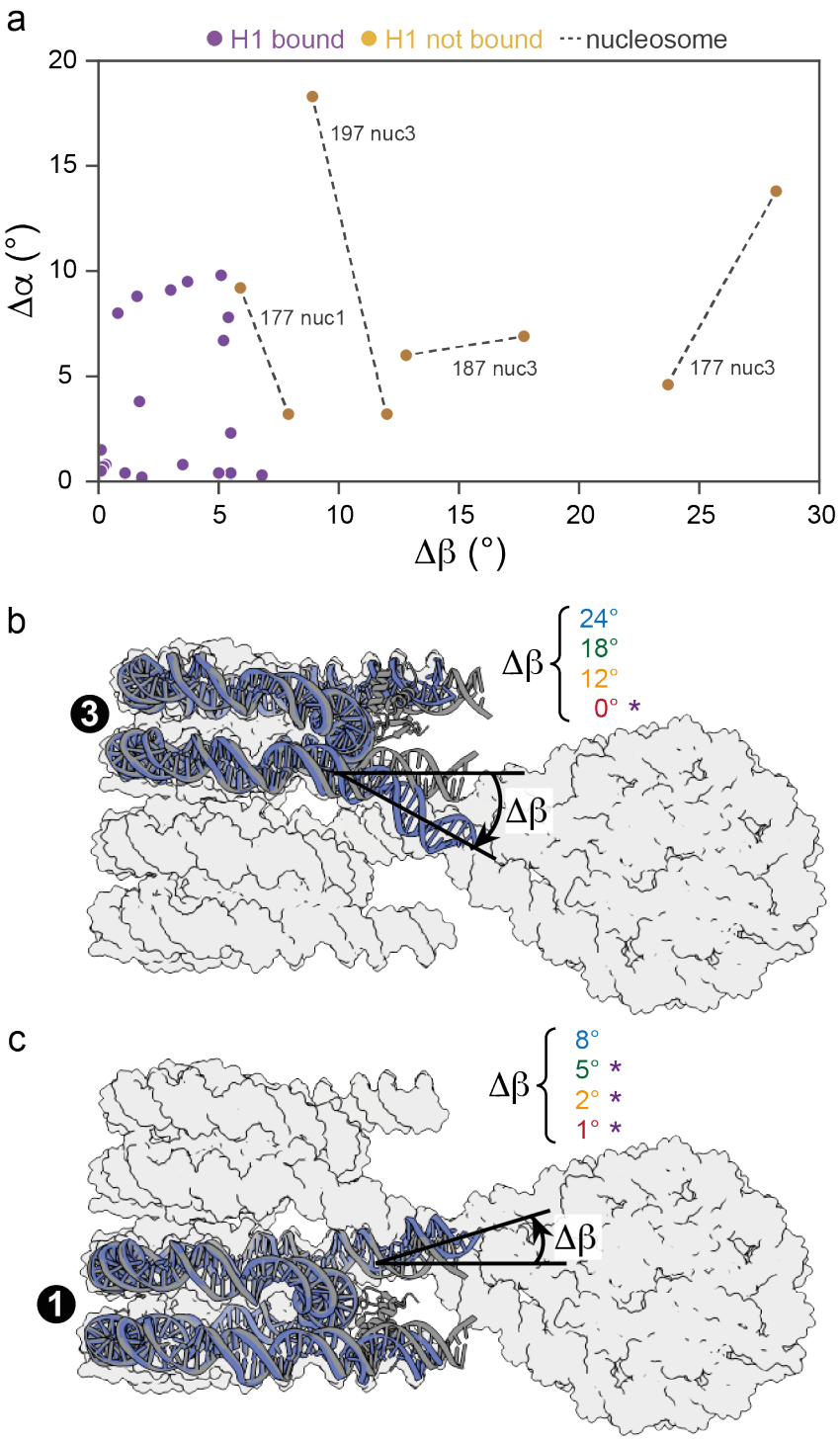

Figure 4.21: *Linker DNA trajectory determines H1 binding. For each nucleosome, $\Delta\alpha$ and $\Delta\beta$ describe the difference in α and β, respectively, between isolated H1-bound mononucleosomal linker DNA (PDB ID 7K5Y (Zhou et al., 2021a)) and the linker DNA of the nucleosomes in the tetranucleosome array. **(a)** A plot of nucleosome entry and exit DNA $\Delta\alpha$ against its $\Delta\beta$ reveals that nucleosomes not bound by H1 (ochre) separate well from the population of nucleosomes bound by H1 (purple). For nucleosome 3, they move closer to this population with increasing NRL. **(b)** $\Delta\beta$ for nucleosome 3 entry DNA reveals a decrease with increasing NRL. **(c)** $\Delta\beta$ for nucleosome 1 exit DNA reveals a decrease with increasing NRL. For the depicted nucleosomes, an overlay of the 4x177 nucleosome (blue) and the isolated H1-bound nucleosome (grey) is shown and $\Delta\beta$ for the different NRL arrays is listed with bound H1 indicated by purple asterisks.

Table 4.5: *Exit and entry DNA geometries at nucleosomes in tetranucleosome arrays.

		nucleosome 1 entry/exit (°)	**nucleosome 2** entry/exit (°)	**nucleosome 3** entry/exit (°)	**nucleosome 4** entry/exit (°)
$\Delta\alpha$	4x177	9.2 / 3.2	7.8 / 0.4	4.6 / 13.8	2.3 / 0.5
	4x187	9.5 / 0.4	0.3 / 9.8	6.9 / 6.0	0.4 / 0.4
	4x197	8.8 / 0.2	0.8 / 6.7	3.2 / 18.3	
	4x207	9.1 / 8.0	0.8 / 0.7	1.5 / 3.8	
$\Delta\beta$	4x177	5.9 / 7.9	5.4 / 5.0	23.7 / 28.2	5.5 / 0.1
	4x187	3.7 / 5.5	6.8 / 5.1	17.7 / 12.8	0.0 / 1.1
	4x197	1.6 / 1.8	3.5 / 5.2	12.0 / 8.9	
	4x207	3.0 / 0.8	0.3 / 0.2	0.1 / 1.7	

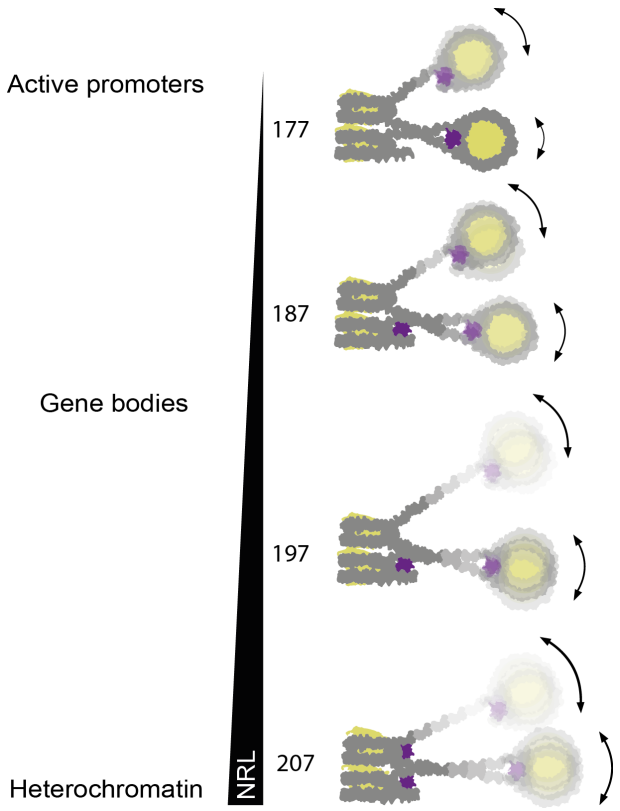

Figure 4.22: *Overview of H1 binding to tetranucleosome arrays. Tetranucleosome arrays bound by linker histone H1 are moderately compacted with nucleosome 1 and nucleosome 3 forming a nucleosome stack. Nucleosome 2 is looped out between the nucleosome stack and does not stack with nucleosome 4. Unstacked nucleosomes have H1 bound. In short NRL arrays found near active promoters, DNA trajectories in stacked nucleosomes deviate and preclude linker histone binding. With increasing NRL, nucleosome 2 moves further away from the nucleosome stack and DNA trajectories in the stacked nucleosomes relax and successively enable linker histone H1 binding to stacked nucleosomes. Long NRLs found in silent heterochromatin enable full linker histone binding to stacked nucleosomes.

Discussion

5.1 The structure of H1 containing nucleosome arrays

The eukaryotic genome is bound by nucleosomes that are arranged on the DNA as 'beads on a string' (Olins and Olins, 1974; Kornberg, 1974; Luger et al., 1997; Oberbeckmann et al., 2019; Baldi et al., 2020). Locally, nucleosomes are regularly spaced and the spacing is related to the transcriptional activity of the underlying genomic sequence and shorter spacings are characteristic of active regions while longer spacings are prevalent in transcriptionally repressed regions (Valouev et al., 2011; Ocampo et al., 2016; Chereji et al., 2018; Lai et al., 2018; Baldi et al., 2018b). Several studies have solved medium resolution structures of *in vitro* reconstituted nucleosome arrays using techniques like x-ray crystallography or cryo-electron microscopy (cryo-EM) of chemically crosslinked samples that decrease inherent array dynamics and have found well ordered and compacted chromatin fibers (Schalch et al., 2005; Song et al., 2014; Ekundayo et al., 2017; Garcia-Saez et al., 2018; Adhireksan et al., 2020). However, *in vivo* nucleosomes rather exist in small disordered clusters (Eltsov et al., 2008; Nishino et al., 2012; Ricci et al., 2015; Chen et al., 2016; Ou et al., 2017; Risca et al., 2017; Cai et al., 2018; Ohno et al., 2019; Xu et al., 2021; Beel et al., 2021). In addition, a relationship between linker histone H1 content and nucleosome repeat length (NRL) has been established (Blank and Becker, 1995; Gunjan et al., 1999; Fan et al., 2005; Woodcock et al., 2006; Eggers and Becker, 2021). While H1 binding to single nucleosomes has been studied at high resolution for different H1 subtypes, how nucleosome arrays are recognized by H1 and the relationship to NRL remains controversial (Zhou et al., 2013; Song et al., 2014; Zhou et al., 2015, 2016; Bednar et al., 2017; Garcia-Saez et al., 2018; Zhou et al., 2018, 2021a). The precise structural nature of short nucleosome arrays with different spacings and its relationship to H1 binding thus remains elusive.

This thesis presents the cryo-electron microscopy (cryo-EM) structures of linker histone H1 containing arrays of four nucleosomes with NRL similar to those found near active promoters (4x177), in gene bodies of actively transcribed genes (4x187, 4x197) and in transcriptionally repressed heterochromatin (4x207) (Valouev et al., 2011). All four structures contain a zig-zag arrangement of nucleosomes and a trinucleosome core

with stacked nucleosomes 1 and 3 and looped out nucleosome 2 that does not stack with nucleosome 4 (Fig. 4.7, 4.16). H1 appears to bind on-dyad and contacts nucleosomal DNA on the dyad and the entry and exit linker DNA (Fig. 4.18a). H1 is bound to all non-stacked nucleosomes but occupancy in stacked nucleosomes increases with NRL (Fig. 4.18b). With increasing NRL, the exit and entry DNA trajectories of stacked nucleosomes change, approach that of the H1-bound isolated mononucleosome and likely enable H1 binding (Fig. 4.20, 4.21, 4.22; Table 4.5). These findings have important implications for understanding chromatin structure and H1 binding in the context of transcriptional activity of nucleosome arrays.

5.1.1 A zig-zag arrangement of nucleosomes

The structures presented here show an arrangement of the nucleosome array in which the linker DNA travels back and forth between consecutive nucleosomes (Figs. 4.7, 4.16). This arrangement is termed zig-zag arrangement or two-start helix (Thoma et al., 1979; Woodcock et al., 1984). Interaction studies *in vitro* and *in vivo* have shown this to be the prevalent local organization of nucleosomes (Dorigo et al., 2004; Risca et al., 2017; Ohno et al., 2019). Similarly, all *in vitro* x-ray crystallography and cryo-EM studies and *in vivo* cryo-electron tomography studies in which the DNA path could confidently be identified show a zig-zag arrangement of nucleosomes (Schalch et al., 2005; Song et al., 2014; Ekundayo et al., 2017; Garcia-Saez et al., 2018; Beel et al., 2021).

Additionally, an interdigitated arrangement of the chromatin fiber called solenoid has been proposed in which nucleosomes are helically arranged along the fiber axis (Finch and Klug, 1976). Simulations suggest that this model might only be viable for very long NRL (Perišić et al., 2010) beyond the 207 bp used here. The 4x207 array adopts a zig-zag confirmation here (Fig. 4.16), but as only a short array of four nucleosomes was used it might be possible for a subpopulation of particles to adopt the solenoid in arrays longer than four nucleosomes. However, there are indications that H3K9me3 chromatin characteristic for HP1 compacted heterochromatin, which has long NRL (Valouev et al., 2011), adapts a zig-zag confirmation *in vivo* (Risca et al., 2017). In summary, the structures of nucleosome arrays across a range of physiologically relevant NRL presented here show a zig-zag arrangement of nucleosomes which has been suggested to be a main component of chromatin architecture (Dorigo et al., 2004; Perišić et al., 2010; Risca et al., 2017; Ohno et al., 2019).

5.1.2 Nucleosome stacking

In all structures of arrays presented here, we observe stacked nucleosomes (Fig. 4.16). Cryo-EM single particle analysis relies on averaging of many particle images and thus biases analyses to populations that cluster very close together in conformational space. This would suggest that stacked nucleosomes frequently occur in nucleosome arrays. Indeed, simulations and small angle x-ray scattering (SAXS) experiments corroborate this (Ding et al., 2021; Mauney et al., 2021). Alas, the relative abundance of stacked nucleosomes over non-stacked nucleosomes cannot be easily estimated by single particle analysis as arrangements with less defined interactions between nucleosomes are likely to be more flexible and cannot be resolved by averaging. However, H1 binding to arrays seems to shift array conformation to stacked nucleosomes, as *in vitro*, nucleosome arrays lacking H1 at similar buffer conditions to the ones used here show an extended configuration of nucleosomes (Takizawa et al., 2020). Overall, stacked nucleosomes thus appear to be a common motif in H1 bound chromatin arrays.

At high resolution, two types of stacking have previously been observed. In type II stacking, nucleosomes are slightly offset; this stacking has been observed in the 6x187 crystal structure (Garcia-Saez et al., 2018) and between tetranucleosome units of the 12x177 and 12x187 (Song et al., 2014). Interactions between the H4 N-terminal tail and the H2A-H2B acidic patch of apposing nucleosomes have been suggested, similar to the interparticle interactions in the mononucleosome crystal (Song et al., 2014; Garcia-Saez et al., 2018; Luger et al., 1997). Type I stacking is very tight with H2A-H2B dimers of apposing nucleosomes being very close together. This stacking has been observed in the 4x167 crystal structure (Schalch et al., 2005) and in the tetranucleosome unit of the 12x177 and 12x187 (Song et al., 2014). In this stacking, the interaction between the H4 N-terminal tail and the acidic patch are not possible (Schalch et al., 2005).

The structures presented here show stacking very similar to type I (Fig. 4.17). Interestingly, this stacking is similar to the stacking observed by cryo-EM in trinucleosomes compacted with physiological Mg^{2+} concentration (Takizawa et al., 2020). As indicated above, H1 binding to the array likely induces this stacking in which the H4 N-terminal tail cannot engage with the acidic patch of the other stacked nucleosome. However, the H4 tail has been shown to be the first point of contact between nucleosomes and to be involved in chromatin folding (Dorigo et al., 2004; Zhang et al., 2017). In this way, the H4 tail might protrude from the nucleosome stack and be available for interactions with

other nucleosomes (Fig. 4.17). Indeed, H1 bound chromatin has been shown to favor interactions between chromatin arrays (Kan et al., 2009). It is thus tempting to speculate that H1 may favor compaction of chromatin by inducing short range interactions such as stacking and long range interactions by making the H4 tail available for interactions with nucleosomes that are further away on the linear genome.

5.1.3 Structural units of chromatin

The cryo-EM structures of the H1 containing 12x177 and 12x187 arrays revealed a twisted arrangement of tetranucleosome units, similar to those of the 4x167 crystal structure, stacked on each other (Song et al., 2014). The twist between tetranucleosome units depended on H1 binding to nucleosomes off-dyad and dimerizing between tetranucleosome units (Song et al., 2014). Studies using single-molecule force spectroscopy on similar arrays revealed length transitions that could be explained by tetranucleosome units successively unfolding (Li et al., 2016) and *in vivo* crosslinking studies found two basic folding motifs of tetranucleosomes (Ohno et al., 2019) to provide further evidence for an array of four nucleosomes forming a recurring structure within chromatin. However, the chemical crosslinking reagent used in the cryo-EM study was shown to alter both H1 binding to arrays and the structure of the chromatin array (Zhou et al., 2018). Additionally, the H1 bound 6x187 crystal structure showed a flat array with uniform stacking along the chromatin fiber (Garcia-Saez et al., 2018). It is thus unclear whether a recurring motif exists within chromatin.

The structures presented here all share a common trinucleosome core (Fig. 4.16). Indeed, there is a study that indicates that trinucleosomes could form a structural unit (Ishihara et al., 2021). There, the authors mildly crosslinked chromatin, fragmented it, analyzed it by sedimentation velocity centrifugation and sequencing and found average fragment sizes of 300-500 bp that could correspond to structural units of 2 to 3 nucleosomes. It is unclear whether trinucleosomes present a significant population of chromatin *in vivo*. Cryo-electron tomography (cryo-ET) of frozen hydrated cells failed to find clear enrichment of distinct units beyond the nucleosome (Cai et al., 2018; Beel et al., 2021). It has been suggested that chromatin binding factors compete with nucleosome-nucleosome interactions and modulate chromatin fiber folding leading to a variety of possible structures depending on the function of the underlying chromatin (Kalashnikova et al., 2013a). These interactions likely contribute to the heterogeneous structure of chromatin in the

nucleus but the accurate determination of the local chromatin interactome and epigenetic status in combination with *in situ* structural biology approaches may prove to be technically challenging.

5.1.4 The effect of salt on chromatin structure

Chromatin structure is highly responsive to salt concentration of the surrounding buffer (Garcia-Saez et al., 2018; Zhou et al., 2018; Takizawa et al., 2020). The structures presented here are solved in the absence of salt under the same conditions used for the cryo-EM structure of the 12x177 and 12x187 (Song et al., 2014). At the beginning of the project, NaCl concentrations between 0 mM and 150 mM were used for cryo-EM grid preparation and screened (Fig. A.1). Of these, grids prepared with arrays in 0 mM NaCl showed well separated particles (Fig. A.1A), whereas clustering or aggregation was observed in sample containing NaCl (Fig. A.1B-D). However, during preparation of other samples, similar behavior was also observed for the sample with 0 mM NaCl, indicating that variables beyond immediate control of the experimenter, such as fluctuations in ice thickness and fluctuations in chamber and room humidity and temperature, likely influence particle behavior in the thin liquid film. This lack of reproducibility is a known phenomenon during cryo-EM grid preparation and methods are being developed to increase reproducibility (Weissenberger et al., 2021).

Salt has been shown to compact nucleosome arrays (Garcia-Saez et al., 2018; Zhou et al., 2018) likely due to screening of the negative charge of DNA (Widom, 1986). The 4x177 trinucleosome core presented here is very similar to the H1 compacted tetranucleosome unit of the 4x177 without salt (Fig. A.2A) (Song et al., 2014) and to the 3x177 lacking H1 but compacted with Mg^{2+} (Fig. A.2B) (Takizawa et al., 2020). It is thus likely that with the addition of salt, nucleosome 4 might stack with nucleosome 2 and lead to similarly distorted DNA trajectories as in the stack consisting of nucleosomes 1 and 3 and destabilize H1 binding there. However, studying the structure of H1 bound tetranucleosome arrays in the presence of physiological salt concentrations will fortify the findings presented in this work.

5.1.5 H1 binding to nucleosomes within arrays

While the structure of H1 bound nucleosomes has been intensely studied at high resolution, how H1 binds to nucleosome arrays is unclear. The globular domain of the somatic variants H1.1 to H1.5 is conserved in humans (Fig. 1.3a) and both H1.4 and H1.5 bind to the mononucleosome contacting nucleosomal DNA on the dyad and interacting with both linker DNAs (Zhou et al., 2015; Bednar et al., 2017; Zhou et al., 2021a). Interestingly, in the context of nucleosome arrays H1.4 appeared to be shifted away from the dyad and interacted with only one linker DNA and seemed to dimerize between tetranucleosome units (Song et al., 2014). Further, H1 seemed to bind off-dyad but different from the off-dyad binding observed for *Drosophila melanogaster* H1 (Zhou et al., 2013). However, the authors of this cryo-EM study used chemical crosslinking that was later shown to change H1 binding to mononucleosome and change the structure of H1 containing nucleosome arrays (Zhou et al., 2018). It is thus unclear whether the arrangement of linker DNA changes H1 binding in nucleosome arrays or whether the difference in binding is a consequence of chemical crosslinking. Surprisingly, the H1 containing 6x187 crystal structure did not allow for accurate identification of H1 density (Garcia-Saez et al., 2018) and could not resolve the controversy of how H1 binds to nucleosome arrays.

The structures presented here were solved from tetranucleosome arrays reconstituted with H1 without the use of any crosslinking agents. The cryo-EM maps showed clear secondary structure elements for the extra density near the nucleosome dyad that could thus be confidently identified as H1 (Figs. 4.8–4.15). Where H1 binding is observed, it binds the nucleosome dyad in the previously described "on-dyad" binding mode (Fig. 4.18) (Zhou et al., 2015; Bednar et al., 2017; Zhou et al., 2021a) and not the "off-dyad" modes observed for *Drosophila* H1 (Zhou et al., 2013) or the crosslinked 12x177 and 12x187 (Song et al., 2014). In arrays unperturbed by chemical crosslinking, full length human H1.4 thus binds to nucleosomes within arrays with its globular domain on the nucleosome dyad.

5.1.6 H1 bound to nucleosome stacks

H1 binding to nucleosomes is on-dyad but slightly off center (Fig. 4.18 and Zhou et al. (2015); Bednar et al. (2017); Zhou et al. (2021a)). It is interesting to note that H1 binds to stacking nucleosomes with its center of mass on the outside of the nucleosome stack (Fig.

4.19). Interestingly, this puts H1 in a similar position to that observed in the tetranucleosome units of the 12x177 and 12x187 cryo-EM structures (Song et al., 2014). There, H1 from adjacent tetranucleosome units might be close enough to be crosslinked together, explaining the observed shift compared to the on-dyad binding mode (Zhou et al., 2018).

As stated above, the orientation of H1 on the nucleosome stack exposes residues of the N-terminal tail domain implicated in regulating H1 function on the surfaces of the nucleosome stack where they would be accessible. Interestingly, methylation of one residue there, K26, was suggested to contribute to the recruitment of HP1 to the nucleosome (Daujat et al., 2005). The structure of an H3K9me3 containing dinucleosome has been solved previously and showed that HP1 dimers bridge consecutive nucleosomes and expose the linker DNA (Machida et al., 2018). However, this study lacked H1 and might not reflect the local environment *in vivo*. Based on the data presented here, it is tempting to speculate that HP1 may bind to and bridge H1 containing H3K9me3-nucleosome stacks to contribute to heterochromatin formation. It is likely that a combination of interactions observed in the cryo-EM structure of the H3K9me3 dinucleosome and the interactions proposed above may be observed.

5.1.7 H1 and NRL

While H1 binds to nucleosomes in the canonical on-dyad mode, exit and entry DNA trajectories appear to modulate whether H1 can actually bind to individual nucleosomes of the array. The structures show that nucleosomes of the array that are not stacked to other nucleosomes (nucleosomes 2 and 4) always have H1 bound (Fig. 4.18). There, entry and exit DNA trajectories do not deviate much from the ideal trajectories observed in H1 bound to isolated nucleosomes (Fig. 4.21 and Table 4.5). In contrast to this, DNA trajectories in stacked nucleosomes 1 and 3 of the 4x177 array deviate substantially from the ideal geometry (Fig. 4.21 and Table 4.5). Nucleosomes 1 exit DNA and nucleosome 3 entry DNA are both connected to nucleosome 2 which has H1 bound and this likely determines the DNA trajectories at nucleosomes 1 and 3. In particular, due to the ca. 90° rotation of nucleosome 2 to stacking nucleosomes 1 and 3, H1 binding to nucleosome 2 might restrict the angle β in stacking nucleosomes 1 and 3. This large deviation might displace DNA far enough to weaken H1 binding in the 4x177 array.

Intriguingly, the deviation decreases with increasing NRL, suggesting that the increased linker DNA length might be able to ameliorate the restrictive effect on DNA

trajectories in nucleosomes 1 and 3 by H1 binding to nucleosome 2 (Fig. 4.21 and Table 4.5). The decrease appears to progress successively as the NRL increases by 10 bp. This is immediately appreciable in nucleosome 3 entry DNA (Fig. 4.21b). Thus, nucleosome 2 that has H1 bound, moves further away from the nucleosome stack with increasing NRL (Fig. 4.16) leading to a decreasing deviation in β for stacking nucleosomes to allow for stable H1 binding in stacking nucleosomes in long NRL arrays (Fig. 4.21 and Table 4.5).

The electrophoretic mobility shift assays (EMSA) confirm that all nucleosomes of all arrays are transiently bound by H1 (Fig. 4.6). However, the large linker DNA trajectory deviation in some nucleosomes may destabilize H1 binding. During cryo-EM sample preparation shearing forces generated by blotting away excess liquid on the cryo-EM grid or effects near the air-water interface (Glaeser, 2018) may lead to dissociation of these weakly bound H1 copies.

The structures presented here also provide a basis for explaining the observed phenomenon of an increased H1 content leading to longer NRL (Fan et al., 2005; Woodcock et al., 2006). Short NRL may disfavor H1 binding but an increase in H1 concentration may lead to an increase in H1 binding to short NRL arrays and may widen the nucleosome footprint against which remodelers space (Oberbeckmann et al., 2021b) until the spacing is long enough to accommodate H1 stoichiometrically. Then, remodeling complexes such as ACF may still space H1 bound nucleosomes and increase the NRL (Maier et al., 2008).

A similar mechanism of H1 exclusion from nucleosomes has been demonstrated for nucleosomes containing the H3 variant CENP-A (Roulland et al., 2016; Zhou et al., 2019; Takizawa et al., 2020). CENP-A contains a shorter α3 helix and binds less DNA at the nucleosome entry and exit sites, leading to a more open conformation that does not permit H1 binding to CENP-A nucleosomes (Zhou et al., 2019). However, the structures presented here go beyond the mononucleosome and elucidate how in nucleosome arrays, the physiologically relevant template for nuclear processes, the arrangement of nucleosomes within an array influences H1 binding and how this arrangement changes with NRL to modulate H1 binding.

5.1.8 Short NRL may contribute to H1 eviction

H1 needs to be depleted from promoters for transcription to occur (Millán-Ariño et al., 2014; Shimada et al., 2019). Several mechanisms of H1 eviction have been proposed, including gene specific pathways dependent on core histone acetylation and NAP1-mediated

eviction (Shimada et al., 2019), modulation by post-translational modifications (Misteli et al., 2000; Izzo and Schneider, 2016; Fyodorov et al., 2018), and direct competition with pioneer transcription factors, HMG proteins or PARP1 (Cirillo et al., 1998, 2002; Krishnakumar et al., 2008; Postnikov and Bustin, 2016).

The data presented here suggest an alternative mechanism that may contribute to H1 depletion or the maintenance of H1 depletion in short NRL arrays. The structures of H1 bound nucleosome arrays explain why H1 can bind to stacking nucleosomes of arrays with long NRL but not to stacking nucleosomes of arrays with short NRL. As indicated above, stacking nucleosomes make up a substantial population of nucleosomes in solution (Ding et al., 2021; Mauney et al., 2021) and stacking appears to be induced by H1 binding (Takizawa et al., 2020). H1 binding may thus be disfavored in short NRL arrays. Intriguingly, short NRL are correlated with higher transcriptional activity and H1 depletion *in vivo* (Ocampo et al., 2016; Chereji et al., 2018). Further, *in vitro* reconstituted chromatin with purified core histones, remodelers, histone chaperones and H1 revealed that chromatin assembled by ACF, which sets long NRL, can incorporate H1, while chromatin assembly by CHD1, which sets short NRL, cannot incorporate H1 (Lusser et al., 2005). It is unclear whether this reflects differences in remodeler activity and interactions or whether this is influenced by the effects seen in this work. To test this, arrays with predefined nucleosome spacings such as the ones used here could be reconstituted and the efficiency of H1 incorporation could be probed by EMSA.

5.1.9 Long NRL may contribute to transcriptional repression

At first glance it seems counterintuitive that transcriptionally repressed heterochromatin should have longer NRL (Valouev et al., 2011; Ocampo et al., 2016; Chereji et al., 2018) as this would expose more DNA. However, the data presented here suggest that H1 may bind more stably to long NRL arrays. *In vitro* studies revealed that while H1 compacted chromatin does not prevent activator binding, it does inhibit transcription (Shimada et al., 2019). Accessibility of DNA may thus not be the only factor determining transcriptional activity (Chereji et al., 2019). Indeed, H1 is though to carry out transcriptional repression by a variety of mechanisms that include physical occlusion of DNA sequences by direct binding, chromatin compaction, recruitment of chromatin modifiers associated with transcriptional repression and inhibition of chromatin modifiers associated with transcriptional activity (Fyodorov et al., 2018; Shimada et al., 2019). Long NRL arrays that bind

H1 more stably may thus be enriched for these repressive interactions and contribute to transcriptional repression.

Outlook

This work presents the structures of tetranucleosomes with a variety of physiologically relevant nucleosome repeat lengths. Due to the absence of crosslinking methods, the structures are dynamic, and extensive computational sorting had to be performed to determine the trinucleosome consensus maps. However, these structures provide a basis for subsequent experiments to improve our knowledge of chromatin structure, linker histone H1 biology and the effect of nucleosome arrays on various aspects of transcription.

6.1 Binding behavior of H1 subtypes

The data presented here indicate that H1 binding may depend on linker DNA trajectories that depend on the higher order structure of nucleosome arrays. Recent simulations based on cryo-EM work suggested that different H1 variants may induce different entry and exit DNA dynamics (Zhou et al., 2021a). Taken together with the observation that different variants induce different spacings in *in vivo* titration experiments, this raises the possibility that different subtypes may be able to be accommodated by different DNA trajectories. It would therefore be interesting to structurally study binding of H1 variants to nucleosome arrays of different NRL and evaluate whether the arrangement of nucleosomes deviates, possibly favoring a less compact arrangement, and whether some variants may bind already in shorter NRL.

6.2 Transcription initiation and nucleosome arrays

In yeast and metazoans, an array of phased nucleosomes extends from the nucleosome depleted region (NDR) into the gene body (Fig. 1.2) (Yuan et al., 2005; Lai and Pugh, 2017). Several lines of evidence suggest that the nucleosome organization near the transcription start site (TSS) modulates transcription initiation. Depletion of a general regulatory factor (GRF) in budding yeast led to transcription initiation occurring away from the canonical TSS (Challal et al., 2018). Concomitant with this ectopic initiation, the au-

thors of this study observed a change in nucleosome positioning (Challal et al., 2018). In fission yeast, CHD1-type remodelers generate the phased array near the TSS and their depletion led to dysregulation of gene expression for some genes (Hennig et al., 2012; Pointner et al., 2012). Additionally, in budding yeast the remodeler RSC was shown to be important for NDR generation and nucleosomes invade the NDR upon RSC depletion, leading to different choices of TSS selections (Klein-Brill et al., 2019).

But while there appears to be a link between promoter proximal chromatin organization and transcription initiation, its precise mechanisms remain elusive. Recent work elucidated the structures of mammalian pre-initiation complexes with the coactivator complex Mediator (Abdella et al., 2021; Chen et al., 2021; Rengachari et al., 2021). However, it remains unclear how the phased array of nucleosomes that is present at the TSS interacts with the PIC and modulates its function. Structural and complementary biochemical studies of transcription initiation upstream of nucleosome arrays will therefore provide insight into how transcription is regulated by its natural chromatin template and how this may be modulated by incorporation of histone variants and by histone post-translational modifications.

6.3 Nucleosome arrays beyond 10n

A meta-analysis of average linker DNA lengths across different cell types and organisms found a quantization of the nucleosome repeat length (NRL) by integer multiples of 10 ($10n$, $n \in \mathbb{N}$) that corresponds roughly to the helical repeat of the DNA double helix (Franklin and Gosling, 1953; Watson and Crick, 1953; Widom, 1992). All published high resolution structures of nucleosome arrays and the ones presented in this work use 10n spacing (Schalch et al., 2005; Song et al., 2014; Ekundayo et al., 2017; Garcia-Saez et al., 2018).

Traditionally, light digestion of chromatin with micrococcal nuclease was used to generate a ladder pattern in agarose gel electrophoresis from which the NRL can be read (Baldi et al., 2020). However, in contrast to yeast, measuring nucleosome positions in metazoan cells is technically challenging as nucleosomes there are poorly positioned (Valouev et al., 2011; Teif et al., 2012; Baldi et al., 2018b). More recently, high throughput next-generation sequencing has been used in combination with various chromatin fragmenting or labeling approaches to determine the NRL more accurately and at individual

loci (Baldi et al., 2018b, 2020). While recent work in yeast suggests that the preferred linker quantization there might be 10n+5, it remains unclear whether preferred quantization exists in metazoans (Wang et al., 2008; Chereji et al., 2018).

Each additional base pair would introduce a circa 36° right-handed rotation leading to a relative rotation of the two sides of a zig-zag array. Indeed, for a trinucleosome with 10n+2 linkers (22 bp) this rotation was observed between the nucleosome stack and the connecting nucleosome (Takizawa et al., 2020). Interestingly, the authors of this study observed both an inward and outward DNA path. The inward DNA path corresponds to the path observed in this work and in previous studies using 10n linkers (Schalch et al., 2005; Song et al., 2014; Ekundayo et al., 2017; Garcia-Saez et al., 2018). It remains unclear whether the fraction of particles with outwards paths changes in different $10n+x$ with $n, x \in \mathbb{N}$ and $0 \leq \mathrm{x} \leq 9$. Simulations suggest different topologies for 10n and 10n+5 fibers and conversion between the two topological states likely necessitates topoisomerase activity (Zhurkin and Norouzi, 2021). There are indications that 10n+5 fibers favor less compact chromatin and might thus be more transcriptionally active in yeast (Zhurkin and Norouzi, 2021). It is tempting to speculate that the effect of increased H1 binding to longer NRL arrays would also be observed there as with increasing linker length, nucleosomes would move further apart and have less of a restrictive effect on the trajectories of neighboring nucleosomes. It would therefore be interesting to structurally characterize H1 binding to 10n+5 arrays and study how the outwards DNA path influences H1 binding in the context of different NRL.

6.4 Chromatin organization at the nuclear lamina

Chromatin at the nuclear lamina is associated with repressed genes and the repressive histone mark H3K9me3 and many multivalent interactions between chromatin and various proteins of the nuclear lamina have been proposed (van Steensel and Belmont, 2017). Among those, HP1 has been suggested to tether chromatin to the nuclear lamina by interacting with the membrane integral lamin B receptor (LBR) (Ye and Worman, 1996) but it is unclear how these interactions work. As indicated above, the structure of the HP1 bound H3K9me3 containing dinucleosome (Machida et al., 2018) might not reflect the local heterochromatin environment *in vivo*. Structural studies of short nucleosome arrays, modified with H3K9me3, bound by H1 trimethylated at K26 in the presence of

HP1 and parts of LBR will be informative for our understanding of chromatin organization near the nuclear lamina.

6.5 The conformational landscape of nucleosome arrays

The structures presented here demonstrate how the three dimensional arrangement of nucleosomes can influence binding of chromatin factors and how chromatin binding factors may affect the structure of chromatin. It is thus likely that an intricate interplay exists that locally leads to favored and disfavored combinations of chromatin binding factors to determine chromatin function. Experimentally probing the conformational landscapes of nucleosome arrays of different NRL with different epigenetic modifications and different chromatin binding factors will thus improve our understanding of chromatin function.

6.5.1 Technical limitations of conventional single particle analysis

The dynamic nature of chromatin is well established in the literature (Garcia-Saez et al., 2018; Zhou et al., 2018). Conventional single particle analysis relies on averaging signal from a large set of particle images (Scheres, 2012). The continuous conformational heterogeneity that is likely present in all macromolecular complexes is routinely treated by assigning particle images into a user defined number of classes (Scheres, 2016) likely leading to a loss of information about internal motion that may be central to macromolecular function.

6.5.2 Probing the conformational landscape of nucleosome arrays

Several approaches to study the conformational landscapes of protein complexes by cryo-electron microscopy have been described. These include but are not limited to three-dimensional (3D) principal component analysis (PCA) (Haselbach et al., 2018), manifold embedding (Frank and Ourmazd, 2016) and machine learning based algorithms (Zhong et al., 2021; Punjani and Fleet, 2021). It remains to be tested whether these approaches are already refined enough to study arrays of nucleosomes, which have too vast a conformational space to yield even medium resolution consensus maps without extensive sorting

of particle images that may lead to loss of information of the conformational landscape.

Beside the computational approaches outlined above, the conformational landscape of chromatin can also conveivably be studied by cryo-electron tomography (cryo-ET). Chromatin exists as repeating units of nucleosomes (Kornberg, 1974). Nucleosomes connected by linker DNA can adapt a variety of relative translations and rotations. Using cryo-ET, three dimensional information on individual particles can be recorded and analyzed (Turk and Baumeister, 2020). In this way, the relative orientations and translations between individual nucleosomes of individual tetranucleosome particles can be determined. Based on these data, PCA based or machine learning based approaches can be used to approximate the conformational landscape of tetranucleosomes. Along these lines, data from simulations of tetranucleosomes has been used to determine the pairwise distances between nucleosome centers in tetranucleosome particles from which a neural net was trained to derive a folding pathway (Ding et al., 2021). It would be interesting to study the effect of different NRL or histone PTMs on the conformational space of nucleosome arrays as there might be combinations of variables that modulate conformational space in a way that could stabilize binding of other factors.

6.5.3 Chromatin organization inside the nucleus

Inside the nucleus, chromatin likely exists with various histone modifications and variants and in complex with a variety of chromatin binding proteins that are locally enriched depending on the genomic region and its transcriptional state. The interactions between chromatin binding proteins have been proposed to shape chromatin structure (Kalashnikova et al., 2013a), making chromatin structure *in* or *ex vivo* very heterogeneous (Cai et al., 2018; Beel et al., 2021; Xu et al., 2021). Nevertheless, the explorative structural study of the nucleus especially in connection with correlative light and electron microscopy (Dobbie, 2019) and focused ion beam milling approaches (Arnold et al., 2016) to target specific genomic loci by fluorescently tagged deactivated Cas9 (Chen et al., 2013; Ma et al., 2015) will substantially improve our understanding of native chromatin structure. Data about the conformational space of nucleosome arrays will help to interpret the arrangement of nucleosomes found *in vivo*.

Bibliography

R. Abdella, A. Talyzina, S. Chen, C. J. Inouye, R. Tjian, and Y. He. Structure of the human Mediator-bound transcription preinitiation complex. *Science*, 372(6537):52–56, 04 2021.

Z. Adhireksan, D. Sharma, P. L. Lee, and C. A. Davey. Near-atomic resolution structures of interdigitated nucleosome fibres. *Nat Commun*, 11(1):4747, 09 2020.

C. Aguilar-Gurrieri, A. Larabi, V. Vinayachandran, N. A. Patel, K. Yen, R. Reja, I. O. Ebong, G. Schoehn, C. V. Robinson, B. F. Pugh, and D. Panne. Structural evidence for Nap1-dependent H2A-H2B deposition and nucleosome assembly. *EMBO J*, 35(13): 1465–1482, 07 2016.

S. Aibara, S. Schilbach, and P. Cramer. Structures of mammalian RNA polymerase II pre-initiation complexes. *Nature*, 594(7861):124–128, 06 2021.

W. Albig and D. Doenecke. The human histone gene cluster at the D6S105 locus. *Hum Genet*, 101(3):284–294, Dec 1997.

W. Albig, B. Bramlage, K. Gruber, H. G. Klobeck, J. Kunz, and D. Doenecke. The human replacement histone H3.3B gene (H3F3B). *Genomics*, 30(2):264–272, Nov 1995.

A. Allahverdi, R. Yang, N. Korolev, Y. Fan, C. A. Davey, C. F. Liu, and L. Nordenskiöld. The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res*, 39(5):1680–1691, Mar 2011.

V. G. Allfrey, R. Faulkner, and A. E. Mirsky. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A*, 51:786–794, May 1964.

C. D. Allis and T. Jenuwein. The molecular hallmarks of epigenetic control. *Nat Rev Genet*, 17(8):487–500, 08 2016.

D. Angelov, J. M. Vitolo, V. Mutskov, S. Dimitrov, and J. J. Hayes. Preferential interaction of the core histone tail domains with linker DNA. *Proc Natl Acad Sci U S A*, 98 (12):6599–6604, Jun 2001.

G. Arents, R. W. Burlingame, B. C. Wang, W. E. Love, and E. N. Moudrianakis. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A*, 88(22):10148–10152, Nov 1991.

J. Arnold, J. Mahamid, V. Lucic, A. de Marco, J. J. Fernandez, T. Laugks, T. Mayer, A. A. Hyman, W. Baumeister, and J. M. Plitzko. Site-Specific Cryo-focused Ion Beam Sample Preparation Guided by 3D Correlative Microscopy. *Biophys J*, 110(4):860–869, Feb 2016.

O. T. Avery, C. M. Macleod, and M. McCarty. Studies on the chemical nature of the substance inducting transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. *J Exp Med*, 79(2):137–158, Feb 1944.

S. Baldi, D. S. Jain, L. Harpprecht, A. Zabel, M. Scheibe, F. Butter, T. Straub, and P. B. Becker. Genome-wide Rules of Nucleosome Phasing in Drosophila. *Mol Cell*, 72(4): 661–672, 11 2018a.

S. Baldi, S. Krebs, H. Blum, and P. B. Becker. Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. *Nat Struct Mol Biol*, 25(9):894–901, 09 2018b.

S. Baldi, P. Korber, and P. B. Becker. Beads on a string-nucleosome array arrangements and folding of the chromatin fiber. *Nat Struct Mol Biol*, 27(2):109–118, 02 2020.

T. Barnes and P. Korber. The Active Mechanism of Nucleosome Depletion by Poly(dA:dT) Tracts In Vivo. *Int J Mol Sci*, 22(15), Jul 2021.

A. D. Baxevanis and D. Landsman. Histone Sequence Database: new histone fold family members. *Nucleic Acids Res*, 26(1):372–375, Jan 1998.

A. D. Baxevanis, J. E. Godfrey, and E. N. Moudrianakis. Associative behavior of the histone (H3-H4)2 tetramer: dependence on ionic environment. *Biochemistry*, 30(36): 8817–8823, Sep 1991.

P. B. Becker and J. L. Workman. Nucleosome remodeling and epigenetics. *Cold Spring Harb Perspect Biol*, 5(9), Sep 2013.

J. Bednar, R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Koster, and C. L. Woodcock. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*, 95(24):14173–14178, Nov 1998.

J. Bednar, I. Garcia-Saez, R. Boopathi, A. R. Cutter, G. Papai, A. Reymer, S. H. Syed, I. N. Lone, O. Tonchev, C. Crucifix, H. Menoni, C. Papin, D. A. Skoufias, H. Kurumizaka, R. Lavery, A. Hamiche, J. J. Hayes, P. Schultz, D. Angelov, C. Petosa, and S. Dimitrov. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell*, 66(3):384–397, May 2017.

A. J. Beel, M. Azubel, P. J. Mattei, and R. D. Kornberg. Structure of mitotic chromosomes. *Mol Cell*, Sep 2021.

S. Bilokapic, M. Strauss, and M. Halic. Cryo-EM of nucleosome core particle interactions in trans. *Sci Rep*, 8(1):7046, 05 2018.

T. A. Blank and P. B. Becker. Electrostatic mechanism of nucleosome spacing. *J Mol Biol*, 252(3):305–313, Sep 1995.

U. Braunschweig, G. J. Hogan, L. Pagie, and B. van Steensel. Histone H1 binding is inhibited by histone variant H3.3. *EMBO J*, 28(23):3635–3645, Dec 2009.

J. E. Brownell, J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth, and C. D. Allis. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, 84(6):843–851, Mar 1996.

S. Cai, D. Böck, M. Pilhofer, and L. Gan. The in situ structures of mono-, di-, and trinucleosomes in human heterochromatin. *Mol Biol Cell*, 29(20):2450–2457, 10 2018.

K. Cao, N. Lailler, Y. Zhang, A. Kumar, K. Uppal, Z. Liu, E. K. Lee, H. Wu, M. Medrzycki, C. Pan, P. Y. Ho, G. P. Cooper, X. Dong, C. Bock, E. E. Bouhassira, and Y. Fan. High-resolution mapping of h1 linker histone variants in embryonic stem cells. *PLoS Genet*, 9(4):e1003417, 2013.

T. L. Caterino and J. J. Hayes. Structure of the H1 C-terminal domain and function in chromatin condensation. *Biochem Cell Biol*, 89(1):35–44, Feb 2011.

D. Challal, M. Barucco, S. Kubik, F. Feuerbach, T. Candelli, H. Geoffroy, C. Benaksas, D. Shore, and D. Libri. General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. *Mol Cell*, 72(6):955–969, 12 2018.

B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, and B. Huang. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7):1479–1491, Dec 2013.

C. Chen, H. H. Lim, J. Shi, S. Tamura, K. Maeshima, U. Surana, and L. Gan. Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. *Mol Biol Cell*, 27(21): 3357–3368, 11 2016.

X. Chen, X. Yin, J. Li, Z. Wu, Y. Qi, X. Wang, W. Liu, and Y. Xu. Structures of the human Mediator and Mediator-bound preinitiation complex. *Science*, 372(6546), 06 2021.

Z. Chen, R. Gabizon, A. I. Brown, A. Lee, A. Song, C. Diaz-Celis, C. D. Kaplan, E. F. Koslover, T. Yao, and C. Bustamante. High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. *Elife*, 8, 07 2019.

R. V. Chereji, S. Ramachandran, T. D. Bryson, and S. Henikoff. Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol*, 19(1):19, 02 2018.

R. V. Chereji, P. R. Eriksson, J. Ocampo, H. K. Prajapati, and D. J. Clark. Accessibility of promoter DNA is not the primary determinant of chromatin-mediated gene regulation. *Genome Res*, 29(12):1985–1995, 12 2019.

C. S. Chu, P. H. Hsu, P. W. Lo, E. Scheer, L. Tora, H. J. Tsai, M. D. Tsai, and L. J. Juan. Protein kinase A-mediated serine 35 phosphorylation dissociates histone H1.4 from mitotic chromosome. *J Biol Chem*, 286(41):35843–35851, Oct 2011.

L. A. Cirillo, C. E. McPherson, P. Bossard, K. Stevens, S. Cherian, E. Y. Shim, K. L. Clark, S. K. Burley, and K. S. Zaret. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J*, 17(1):244–254, Jan 1998.

L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, and K. S. Zaret. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*, 9(2):279–289, Feb 2002.

C. R. Clapier, J. Iwasa, B. R. Cairns, and C. L. Peterson. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*, 18(7):407–422, 07 2017.

R. Dahm and F. Miescher. Friedrich Miescher and the discovery of DNA. *Dev Biol*, 278 (2):274–288, Feb 2005.

S. Daujat, U. Zeissler, T. Waldmann, N. Happel, and R. Schneider. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J Biol Chem*, 280(45):38090–38095, Nov 2005.

C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol*, 319(5):1097–1113, Jun 2002.

M. L. Dechassa, K. Wyns, M. Li, M. A. Hall, M. D. Wang, and K. Luger. Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat Commun*, 2: 313, 2011.

C. Dhalluin, J. E. Carlson, L. Zeng, C. He, A. K. Aggarwal, and M. M. Zhou. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, 399(6735):491–496, Jun 1999.

X. Ding, X. Lin, and B. Zhang. Stability and folding pathways of tetra-nucleosome from six-dimensional free energy surface. *Nat Commun*, 12(1):1091, 02 2021.

I. M. Dobbie. Bridging the resolution gap: correlative super-resolution imaging. *Nat Rev Microbiol*, 17(6):337, 06 2019.

Z. Dominski and W. F. Marzluff. Formation of the 3' end of histone mRNA. *Gene*, 239 (1):1–14, Oct 1999.

B. Dorigo, T. Schalch, K. Bystricky, and T. J. Richmond. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol*, 327(1):85–96, Mar 2003.

B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder, and T. J. Richmond. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306(5701):1571–1573, Nov 2004.

H. R. Drew and A. A. Travers. DNA bending and its relation to nucleosome positioning. *J Mol Biol*, 186(4):773–790, Dec 1985.

I. Dunham, A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Frietze, J. Harrow, R. Kaul, J. Khatun, B. R. Lajoie, S. G. Landt, B. K. Lee, F. Pauli, K. R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shoresh, J. M. Simon, L. Song, N. D. Trinklein, R. C. Altshuler, E. Birney, J. B. Brown, C. Cheng, S. Djebali, X. Dong, I. Dunham, J. Ernst, T. S. Furey, M. Gerstein, B. Giardine, M. Greven, R. C. Hardison, R. S. Harris, J. Herrero, M. M. Hoffman, S. Iyer, M. Kellis, J. Khatun, P. Kheradpour, A. Kundaje, T. Lassmann, Q. Li, X. Lin, G. K. Marinov, A. Merkel, A. Mortazavi, S. C. Parker, T. E. Reddy, J. Rozowsky, F. Schlesinger, R. E. Thurman, J. Wang, L. D. Ward, T. W. Whitfield, S. P. Wilder, W. Wu, H. S. Xi, K. Y. Yip, J. Zhuang, M. J. Pazin, R. F. Lowdon, L. A. Dillon, L. B. Adams, C. J. Kelly, J. Zhang, J. R. Wexler, E. D. Green, P. J. Good, E. A. Feingold, B. E. Bernstein, E. Birney, G. E. Crawford, J. Dekker, L. Elnitski, P. J. Farnham, M. Gerstein, M. C. Giddings, T. R. Gingeras, E. D. Green, R. Guigo, R. C. Hardison, T. J. Hubbard, M. Kellis, W. Kent, J. D. Lieb, E. H. Margulies, R. M. Myers, M. Snyder, J. A. Stamatoyannopoulos, S. A. Tenenbaum, Z. Weng, K. P. White, B. Wold, J. Khatun, Y. Yu, J. Wrobel, B. A. Risk, H. P. Gunawardena, H. C. Kuiper, C. W. Maier, L. Xie, X. Chen, M. C. Giddings, B. E. Bernstein, C. B. Epstein, N. Shoresh, J. Ernst, P. Kheradpour, T. S. Mikkelsen, S. Gillespie, A. Goren, O. Ram, X. Zhang, L. Wang, R. Issner, M. J. Coyne, T. Durham, M. Ku, T. Truong, L. D. Ward, R. C. Altshuler, M. L. Eaton, M. Kellis, S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Roder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, P. Batut, I. Bell, K. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. P. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, G. Li, O. J. Luo, E. Park, J. B. Preall, K. Presaud, P. Ribeca, B. A. Risk, D. Robyr, X. Ruan, M. Sammeth, K. S. Sandhu, L. Schaeffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, Y. Hayashizaki, J. Harrow, M. Gerstein, T. J. Hubbard, A. Reymond, S. E. Antonarakis, G. J. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigo, T. R. Gingeras, K. R. Rosenbloom, C. A. Sloan, K. Learned, V. S. Malladi, M. C. Wong, G. P. Barber, M. S. Cline, T. R. Dreszer, S. G. Heitner, D. Karolchik, W. Kent, V. M. Kirkup, L. R. Meyer, J. C. Long, M. Maddren, B. J. Raney, T. S. Furey, L. Song, L. L. Grasfeder, P. G. Giresi, B. K. Lee, A. Battenhouse, N. C. Sheffield, J. M. Simon, K. A. Showers, A. Safi, D. London, A. A. Bhinge, C. Shestak, M. R. Schaner, S. K. Kim, Z. Z. Zhang, P. A. Mieczkowski,

J. O. Mieczkowska, Z. Liu, R. M. McDaniell, Y. Ni, N. U. Rashid, M. J. Kim, S. Adar, Z. Zhang, T. Wang, D. Winter, D. Keefe, E. Birney, V. R. Iyer, J. D. Lieb, G. E. Crawford, G. Li, K. S. Sandhu, M. Zheng, P. Wang, O. J. Luo, A. Shahab, M. J. Fullwood, X. Ruan, Y. Ruan, R. M. Myers, F. Pauli, B. A. Williams, J. Gertz, G. K. Marinov, T. E. Reddy, J. Vielmetter, E. Partridge, D. Trout, K. E. Varley, C. Gasper, A. Bansal, S. Pepke, P. Jain, H. Amrhein, K. M. Bowling, M. Anaya, M. K. Cross, B. King, M. A. Muratet, I. Antoshechkin, K. M. Newberry, K. McCue, A. S. Nesmith, K. I. Fisher-Aylor, B. Pusey, G. DeSalvo, S. L. Parker, S. Balasubramanian, N. S. Davis, S. K. Meadows, T. Eggleston, C. Gunter, J. Newberry, S. E. Levy, D. M. Absher, A. Mortazavi, W. H. Wong, B. Wold, M. J. Blow, A. Visel, L. A. Pennachio, L. Elnitski, E. H. Margulies, S. C. Parker, H. M. Petrykowska, A. Abyzov, B. Aken, D. Barrell, G. Barson, A. Berry, A. Bignell, V. Boychenko, G. Bussotti, J. Chrast, C. Davidson, T. Derrien, G. Despacio-Reyes, M. Diekhans, I. Ezkurdia, A. Frankish, J. Gilbert, J. M. Gonzalez, E. Griffiths, R. Harte, D. A. Hendrix, C. Howald, T. Hunt, I. Jungreis, M. Kay, E. Khurana, F. Kokocinski, J. Leng, M. F. Lin, J. Loveland, Z. Lu, D. Manthravadi, M. Mariotti, J. Mudge, G. Mukherjee, C. Notredame, B. Pei, J. M. Rodriguez, G. Saunders, A. Sboner, S. Searle, C. Sisu, C. Snow, C. Steward, A. Tanzer, E. Tapanari, M. L. Tress, M. J. van Baren, N. Walters, S. Washietl, L. Wilming, A. Zadissa, Z. Zhang, M. Brent, D. Haussler, M. Kellis, A. Valencia, M. Gerstein, A. Reymond, R. Guigo, J. Harrow, T. J. Hubbard, S. G. Landt, S. Frietze, A. Abyzov, N. Addleman, R. P. Alexander, R. K. Auerbach, S. Balasubramanian, K. Bettinger, N. Bhardwaj, A. P. Boyle, A. R. Cao, P. Cayting, A. Charos, Y. Cheng, C. Cheng, C. Eastman, G. Euskirchen, J. D. Fleming, F. Grubert, L. Habegger, M. Hariharan, A. Harmanci, S. Iyengar, V. X. Jin, K. J. Karczewski, M. Kasowski, P. Lacroute, H. Lam, N. Lamarre-Vincent, J. Leng, J. Lian, M. Lindahl-Allen, R. Min, B. Miotto, H. Monahan, Z. Moqtaderi, X. J. Mu, H. O'Geen, Z. Ouyang, D. Patacsil, B. Pei, D. Raha, L. Ramirez, B. Reed, J. Rozowsky, A. Sboner, M. Shi, C. Sisu, T. Slifer, H. Witt, L. Wu, X. Xu, K. K. Yan, X. Yang, K. Y. Yip, Z. Zhang, K. Struhl, S. M. Weissman, M. Gerstein, P. J. Farnham, M. Snyder, S. A. Tenenbaum, L. O. Penalva, F. Doyle, S. Karmakar, S. G. Landt, R. R. Bhanvadia, A. Choudhury, M. Domanus, L. Ma, J. Moran, D. Patacsil, T. Slifer, A. Victorsen, X. Yang, M. Snyder, T. Auer, L. Centanin, M. Eichenlaub, F. Gruhl, S. Heermann, B. Hoeckendorf, D. Inoue, T. Kellner, S. Kirchmaier, C. Mueller, R. Reinhardt, L. Schertel, S. Schneider, R. Sinn, B. Wittbrodt, J. Wittbrodt, Z. Weng, T. W. Whitfield, J. Wang, P. J. Collins, S. F. Aldred, N. D. Trinklein, E. C. Partridge, R. M. Myers, J. Dekker, G. Jain, B. R. Lajoie, A. Sanyal, G. Balasundaram, D. L. Bates, R. Byron, T. K. Canfield, M. J. Diegel, D. Dunn, A. K. Ebersol, T. Frum, K. Garg, E. Gist, R. Hansen, L. Boatman, E. Haugen, R. Humbert, G. Jain, A. K. Johnson, E. M. Johnson, T. V. Kutyavin, B. R. Lajoie, K. Lee, D. Lotakis, M. T. Maurano, S. J. Neph, F. V. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, E. Rynes, P. Sabo, M. E. Sanchez, R. S. Sandstrom, A. Sanyal, A. O. Shafer, A. B. Stergachis, S. Thomas, R. E. Thurman, B. Vernot, J. Vierstra, S. Vong, H. Wang, M. A. Weaver, Y. Yan, M. Zhang, J. M. Akey, M. Bender, M. O. Dorschner, M. Groudine, M. J. MacCoss, P. Navas,

G. Stamatoyannopoulos, R. Kaul, J. Dekker, J. A. Stamatoyannopoulos, I. Dunham, K. Beal, A. Brazma, P. Flicek, J. Herrero, N. Johnson, D. Keefe, M. Lukk, N. M. Luscombe, D. Sobral, J. M. Vaquerizas, S. P. Wilder, S. Batzoglou, A. Sidow, N. Hussami, S. Kyriazopoulou-Panagiotopoulou, M. W. Libbrecht, M. A. Schaub, A. Kundaje, R. C. Hardison, W. Miller, B. Giardine, R. S. Harris, W. Wu, P. J. Bickel, B. Banfai, N. P. Boley, J. B. Brown, H. Huang, Q. Li, J. J. Li, W. S. Noble, J. A. Bilmes, O. J. Buske, M. M. Hoffman, A. D. Sahu, P. V. Kharchenko, P. J. Park, D. Baker, J. Taylor, Z. Weng, S. Iyer, X. Dong, M. Greven, X. Lin, J. Wang, H. S. Xi, J. Zhuang, M. Gerstein, R. P. Alexander, S. Balasubramanian, C. Cheng, A. Harmanci, L. Lochovsky, R. Min, X. J. Mu, J. Rozowsky, K. K. Yan, K. Y. Yip, and E. Birney. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, Sep 2012.

P. N. Dyer, R. S. Edayathumangalam, C. L. White, Y. Bao, S. Chakravarthy, U. M. Muthurajan, and K. Luger. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol*, 375:23–44, 2004.

N. Eggers and P. B. Becker. Cell-free genomics reveal intrinsic, cooperative and competitive determinants of chromatin interactions. *Nucleic Acids Res*, 49(13):7602–7617, 07 2021.

B. Ekundayo, T. J. Richmond, and T. Schalch. Capturing Structural Heterogeneity in Chromatin Fibers. *J Mol Biol*, 429(20):3031–3042, 10 2017.

M. Eltsov, K. M. Maclellan, K. Maeshima, A. S. Frangakis, and J. Dubochet. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc Natl Acad Sci U S A*, 105(50):19732–19737, Dec 2008.

Y. Fan, T. Nikitina, J. Zhao, T. J. Fleury, R. Bhattacharyya, E. E. Bouhassira, A. Stein, C. L. Woodcock, and A. I. Skoultchi. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, 123(7):1199–1212, Dec 2005.

L. Farnung, S. M. Vos, and P. Cramer. Structure of transcribing RNA polymerase II-nucleosome complex. *Nat Commun*, 9(1):5432, 12 2018.

L. Farnung, M. Ochmann, M. Engeholm, and P. Cramer. Structural basis of nucleosome transcription mediated by Chd1 and FACT. *Nat Struct Mol Biol*, 28(4):382–387, 04 2021.

J. T. Finch and A. Klug. Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*, 73(6):1897–1901, Jun 1976.

W. Fischle, B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki, and C. D. Allis. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*, 438(7071):1116–1122, Dec 2005.

A. Flaus, D. M. Martin, G. J. Barton, and T. Owen-Hughes. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res*, 34(10): 2887–2905, 2006.

J. Frank and A. Ourmazd. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods*, 100:61–67, 05 2016.

R. E. Franklin and R. G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740–741, Apr 1953.

E. Fussner, R. W. Ching, and D. P. Bazett-Jones. Living without 30nm chromatin fibers. *Trends Biochem Sci*, 36(1):1–6, Jan 2011.

D. V. Fyodorov, B. R. Zhou, A. I. Skoultchi, and Y. Bai. Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol*, 19(3):192–206, 03 2018.

I. Garcia-Saez, H. Menoni, R. Boopathi, M. S. Shukla, L. Soueidan, M. Noirclerc-Savoye, A. Le Roy, D. A. Skoufias, J. Bednar, A. Hamiche, D. Angelov, C. Petosa, and S. Dimitrov. Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. *Mol Cell*, 72(5):902–915, 12 2018.

R. M. Glaeser. PROTEINS, INTERFACES, AND CRYO-EM GRIDS. *Curr Opin Colloid Interface Sci*, 34:1–8, Mar 2018.

O. Glaich, Y. Leader, G. Lev Maor, and G. Ast. Histone H1.5 binds over splice sites in chromatin and regulates alternative splicing. *Nucleic Acids Res*, 47(12):6145–6159, 07 2019.

R. S. Grand, L. Burger, C. Gräwe, A. K. Michael, L. Isbel, D. Hess, L. Hoerner, V. Iesmantavicius, S. Durdu, M. Pregnolato, A. R. Krebs, S. A. Smallwood, N. Thomä, M. Vermeulen, and D. Schübeler. BANP opens chromatin and activates CpG-island-regulated genes. *Nature*, 596(7870):133–137, 08 2021.

S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc Natl Acad Sci U S A*, 106(32):13317–13322, Aug 2009.

F. Guerrero, A. Ciragan, and H. Iwai. Tandem SUMO fusion vectors for improving soluble protein expression and purification. *Protein Expr Purif*, 116:42–49, Dec 2015.

A. Gunjan, B. T. Alexander, D. B. Sittman, and D. T. Brown. Effects of H1 histone variant overexpression on chromatin structure. *J Biol Chem*, 274(53):37950–37956, Dec 1999.

C. M. Hammond, C. B. Strømme, H. Huang, D. J. Patel, and A. Groth. Histone chaperone networks shaping chromatin function. *Nat Rev Mol Cell Biol*, 18(3):141–158, 03 2017.

M. Hantsche and P. Cramer. The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angew Chem Int Ed Engl*, 55(52):15972–15981, 12 2016.

F. Hao, K. J. Murphy, T. Kujirai, N. Kamo, J. Kato, M. Koyama, A. Okamato, G. Hayashi, H. Kurumizaka, and J. J. Hayes. Acetylation-modulated communication between the H3 N-terminal tail domain and the intrinsically disordered H1 C-terminal domain. *Nucleic Acids Res*, 48(20):11510–11520, 11 2020.

J. M. Harp, B. L. Hanson, D. E. Timm, and G. J. Bunick. Asymmetries in the nucleosome core particle at 2.5 A resolution. *Acta Crystallogr D Biol Crystallogr*, 56(Pt 12):1513–1534, Dec 2000.

D. Haselbach, I. Komarov, D. E. Agafonov, K. Hartmuth, B. Graf, O. Dybkov, H. Urlaub, B. Kastner, R. Lührmann, and H. Stark. Complex. *Cell*, 172(3):454–464, 01 2018.

C. L. Hatch and W. M. Bonner. The human histone H2A.Z gene. Sequence and regulation. *J Biol Chem*, 265(25):15211–15218, Sep 1990.

S. E. Healton, H. D. Pinto, L. N. Mishra, G. A. Hamilton, J. C. Wheat, K. Swist-Rosowska, N. Shukeir, Y. Dou, U. Steidl, T. Jenuwein, M. J. Gamble, and A. I. Skoultchi. H1 linker histones silence repetitive elements by promoting both histone H3K9 methylation and chromatin compaction. *Proc Natl Acad Sci U S A*, 117(25):14251–14258, 06 2020.

M. J. Hendzel, M. A. Lever, E. Crawford, and J. P. Th'ng. The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J Biol Chem*, 279 (19):20028–20034, May 2004.

B. P. Hennig, K. Bendrin, Y. Zhou, and T. Fischer. Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO Rep*, 13(11):997–1003, Nov 2012.

S. P. Hergeth and R. Schneider. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep*, 16(11):1439–1453, Nov 2015.

J. E. Herrera, K. L. West, R. L. Schiltz, Y. Nakatani, and M. Bustin. Histone H1 is a specific repressor of core histone acetylation in chromatin. *Mol Cell Biol*, 20(2):523–529, Jan 2000.

C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science*, 325(5940): 626–628, Jul 2009.

K. E. Holde. Chromatin. *Springer Series in Molecular and Cell Biology. Springer-Verlag, New York*, 1998.

S. Ishihara, Y. Sasagawa, T. Kameda, H. Yamashita, M. Umeda, N. Kotomura, M. Abe, Y. Shimono, and I. Nikaido. Local states of chromatin compaction at transcription start sites control transcription levels. *Nucleic Acids Res*, 49(14):8007–8023, 08 2021.

M. G. Izban and D. S. Luse. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem*, 267(19):13647–13655, Jul 1992.

A. Izzo and R. Schneider. The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. *Biochim Biophys Acta*, 1859(3):486–495, Mar 2016.

A. Izzo, K. Kamieniarz, and R. Schneider. The histone H1 family: specific members, specific functions? *Biol Chem*, 389(4):333–343, Apr 2008.

A. Izzo, K. Kamieniarz-Gdula, F. Ramírez, N. Noureen, J. Kind, T. Manke, B. van Steensel, and R. Schneider. The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Rep*, 3(6):2142–2154, Jun 2013.

T. Jenuwein and C. D. Allis. Translating the histone code. *Science*, 293(5532):1074–1080, Aug 2001.

A. A. Kalashnikova, M. E. Porter-Goff, U. M. Muthurajan, K. Luger, and J. C. Hansen. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J R Soc Interface*, 10(82):20121022, May 2013a.

A. A. Kalashnikova, D. D. Winkler, S. J. McBryant, R. K. Henderson, J. A. Herman, J. G. DeLuca, K. Luger, J. E. Prenni, and J. C. Hansen. Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. *Nucleic Acids Res*, 41(7): 4026–4035, Apr 2013b.

K. Kamieniarz, A. Izzo, M. Dundr, P. Tropberger, L. Ozretic, J. Kirfel, E. Scheer, P. Tropel, J. R. Wisniewski, L. Tora, S. Viville, R. Buettner, and R. Schneider. A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes Dev*, 26 (8):797–802, Apr 2012.

P. Y. Kan, T. L. Caterino, and J. J. Hayes. The H4 tail domain participates in intra- and internucleosome interactions with protein and DNA during folding and oligomerization of nucleosome arrays. *Mol Cell Biol*, 29(2):538–546, Jan 2009.

R. I. Kelley. Isolation of a histone IIb1-IIb2 complex. *Biochem Biophys Res Commun*, 54 (4):1588–1594, Oct 1973.

K. Kim, B. Lee, J. Kim, J. Choi, J. M. Kim, Y. Xiong, R. G. Roeder, and W. An. Linker Histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation. *Cell Rep*, 5(6):1690–1703, Dec 2013.

A. Klein-Brill, D. Joseph-Strauss, A. Appleboim, and N. Friedman. Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. *Cell Rep*, 26(1):279–292, 01 2019.

S. L. Klemm, Z. Shipony, and W. J. Greenleaf. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*, 20(4):207–220, 04 2019.

R. D. Kornberg. Chromatin structure: a repeating unit of histones and DNA. *Science*, 184(4139):868–871, May 1974.

R. D. Kornberg and J. O. Thomas. Chromatin structure; oligomers of the histones. *Science*, 184(4139):865–868, May 1974.

A. Kossel. über einen peptonartigen bestandtheil des zellkerns. *Zeitschrift für Physiologische Chemie*, 8:511, 1884.

N. Krietenstein, M. Wal, S. Watanabe, B. Park, C. L. Peterson, B. F. Pugh, and P. Korber. Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell*, 167(3):709–721, Oct 2016.

R. Krishnakumar, M. J. Gamble, K. M. Frizzell, J. G. Berrocal, M. Kininis, and W. L. Kraus. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science*, 319(5864):819–821, Feb 2008.

T. Kujirai, H. Ehara, Y. Fujino, M. Shirouzu, S. I. Sekine, and H. Kurumizaka. Structural basis of the nucleosome transition during RNA polymerase II passage. *Science*, 362 (6414):595–598, 11 2018.

M. H. Kuo, J. E. Brownell, R. E. Sobel, T. A. Ranalli, R. G. Cook, D. G. Edmondson, S. Y. Roth, and C. D. Allis. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature*, 383(6597):269–272, Sep 1996.

B. Lai, W. Gao, K. Cui, W. Xie, Q. Tang, W. Jin, G. Hu, B. Ni, and K. Zhao. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature*, 562(7726):281–285, 10 2018.

W. K. M. Lai and B. F. Pugh. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol*, 18(9):548–562, Sep 2017.

R. A. Laskey, B. M. Honda, A. D. Mills, and J. T. Finch. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature*, 275(5679): 416–420, Oct 1978.

P. J. Laybourn and J. T. Kadonaga. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science*, 254(5029):238–245, Oct 1991.

M. A. Lever, J. P. Th'ng, X. Sun, and M. J. Hendzel. Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, 408(6814):873–876, Dec 2000.

B. Li, M. Carey, and J. L. Workman. The role of chromatin during transcription. *Cell*, 128(4):707–719, Feb 2007.

W. Li, P. Chen, J. Yu, L. Dong, D. Liang, J. Feng, J. Yan, P. Y. Wang, Q. Li, Z. Zhang, M. Li, and G. Li. FACT Remodels the Tetranucleosomal Unit of Chromatin Fibers for Gene Transcription. *Mol Cell*, 64(1):120–133, 10 2016.

C. Lieleg, N. Krietenstein, M. Walker, and P. Korber. Nucleosome positioning in yeasts: methods, maps, and mechanisms. *Chromosoma*, 124(2):131–151, Jun 2015.

J. T. Lis and R. Schleif. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic Acids Res*, 2(3):383–389, Mar 1975.

Y. Lorch and R. D. Kornberg. Chromatin-remodeling for transcription. *Q Rev Biophys*, 50:e5, 01 2017.

Y. Lorch, J. W. LaPointe, and R. D. Kornberg. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, 49(2): 203–210, Apr 1987.

P. T. Lowary and J. Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*, 276(1):19–42, Feb 1998.

X. Lu, S. N. Wontakal, H. Kavi, B. J. Kim, P. M. Guzzardo, A. V. Emelyanov, N. Xu, G. J. Hannon, J. Zavadil, D. V. Fyodorov, and A. I. Skoultchi. Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. *Science*, 340 (6128):78–81, Apr 2013.

K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251–260, Sep 1997.

K. Luger, T. J. Rechsteiner, and T. J. Richmond. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol*, 119:1–16, 1999.

A. Lusser, D. L. Urwin, and J. T. Kadonaga. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol*, 12(2):160–166, Feb 2005.

H. Ma, A. Naseri, P. Reyes-Gutierrez, S. A. Wolfe, S. Zhang, and T. Pederson. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A*, 112 (10):3002–3007, Mar 2015.

S. Machida, Y. Takizawa, M. Ishimaru, Y. Sugita, S. Sekine, J. I. Nakayama, M. Wolf, and H. Kurumizaka. Structural Basis of Heterochromatin Formation by Human HP1. *Mol Cell*, 69(3):385–397, 02 2018.

K. Maeshima, S. Hihara, and M. Eltsov. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol*, 22(3):291–297, Jun 2010.

K. Maeshima, R. Imai, S. Tamura, and T. Nozaki. Chromatin as dynamic 10-nm fibers. *Chromosoma*, 123(3):225–237, Jun 2014.

V. K. Maier, M. Chioda, D. Rhodes, and P. B. Becker. ACF catalyses chromatosome movements in chromatin fibres. *EMBO J*, 27(6):817–826, Mar 2008.

W. F. Marzluff, P. Gongidi, K. R. Woods, J. Jin, and L. J. Maltais. The human and mouse replication-dependent histone genes. *Genomics*, 80(5):487–498, Nov 2002.

A. W. Mauney, U. M. Muthurajan, K. Luger, and L. Pollack. Solution structure(s) of trinucleosomes from contrast variation SAXS. *Nucleic Acids Res*, 49(9):5028–5037, 05 2021.

R. Mayor, A. Izquierdo-Bouldstridge, L. Millán-Ariño, A. Bustillos, C. Sampaio, N. Luque, and A. Jordan. Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions. *J Biol Chem*, 290(12):7474–7491, Mar 2015.

R. K. McGinty and S. Tan. Nucleosome structure and function. *Chem Rev*, 115(6): 2255–2273, Mar 2015.

R. K. McGinty and S. Tan. Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol*, 37:54–61, Apr 2016.

R. K. McGinty and S. Tan. Principles of nucleosome recognition by chromatin factors and enzymes. *Curr Opin Struct Biol*, 71:16–26, Jun 2021.

L. Millán-Ariño, A. B. Islam, A. Izquierdo-Bouldstridge, R. Mayor, J. M. Terme, N. Luque, M. Sancho, N. López-Bigas, and A. Jordan. Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. *Nucleic Acids Res*, 42(7):4474–4493, Apr 2014.

L. Millán-Ariño, A. Izquierdo-Bouldstridge, and A. Jordan. Specificities and genomic distribution of somatic mammalian histone H1 subtypes. *Biochim Biophys Acta*, 1859 (3):510–519, Mar 2016.

T. Misteli, A. Gunjan, R. Hock, M. Bustin, and D. T. Brown. Dynamic binding of histone H1 to chromatin in living cells. *Nature*, 408(6814):877–881, Dec 2000.

C. A. Mizzen, X. J. Yang, T. Kokubo, J. E. Brownell, A. J. Bannister, T. Owen-Hughes, J. Workman, L. Wang, S. L. Berger, T. Kouzarides, Y. Nakatani, and C. D. Allis. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell*, 87(7): 1261–1270, Dec 1996.

E. A. Morrison, S. Bowerman, K. L. Sylvers, J. Wereszczynski, and C. A. Musselman. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife*, 7, 04 2018.

H. C. Nelson, J. T. Finch, B. F. Luisi, and A. Klug. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature*, 330(6145):221–226, 1987.

Y. Nishino, M. Eltsov, Y. Joti, K. Ito, H. Takata, Y. Takahashi, S. Hihara, A. S. Frangakis, N. Imamoto, T. Ishikawa, and K. Maeshima. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J*, 31(7):1644–1653, Apr 2012.

E. Oberbeckmann, M. Wolff, N. Krietenstein, M. Heron, J. L. Ellins, A. Schmid, S. Krebs, H. Blum, U. Gerland, and P. Korber. genome. *Genome Res*, 29(12):1996–2009, 12 2019.

E. Oberbeckmann, N. Krietenstein, V. Niebauer, Y. Wang, K. Schall, M. Moldt, T. Straub, R. Rohs, K. P. Hopfner, P. Korber, and S. Eustermann. Genome information processing by the INO80 chromatin remodeler positions nucleosomes. *Nat Commun*, 12(1):3231, 05 2021a.

E. Oberbeckmann, V. Niebauer, S. Watanabe, L. Farnung, M. Moldt, A. Schmid, P. Cramer, C. L. Peterson, S. Eustermann, K. P. Hopfner, and P. Korber. Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. *Nat Commun*, 12(1):3232, 05 2021b.

C. Öberg, A. Izzo, R. Schneider, Ö. Wrange, and S. Belikov. Linker histone subtypes differ in their effect on nucleosomal spacing in vivo. *J Mol Biol*, 419(3-4):183–197, Jun 2012.

J. Ocampo, R. V. Chereji, P. R. Eriksson, and D. J. Clark. The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. *Nucleic Acids Res*, 44(10):4625–4635, 06 2016.

V. V. Ogryzko, R. L. Schiltz, V. Russanova, B. H. Howard, and Y. Nakatani. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, 87(5): 953–959, Nov 1996.

M. Ohno, T. Ando, D. G. Priest, V. Kumar, Y. Yoshida, and Y. Taniguchi. Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs. *Cell*, 176 (3):520–534, 01 2019.

A. L. Olins and D. E. Olins. Spheroid chromatin units (v bodies). *Science*, 183(4122): 330–332, Jan 1974.

A. Osunsade, N. A. Prescott, J. M. Hebert, D. M. Ray, Y. Jmeian, I. C. Lorenz, and Y. David. A Robust Method for the Purification and Characterization of Recombinant Human Histone H1 Variants. *Biochemistry*, 58(3):171–176, 01 2019.

H. D. Ou, S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'Shea. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349), 07 2017.

N. B. Pandey, N. Chodchoy, T. J. Liu, and W. F. Marzluff. Introns in histone genes alter the distribution of 3' ends. *Nucleic Acids Res*, 18(11):3161–3170, Jun 1990.

J. Pei, B. H. Kim, and N. V. Grishin. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res*, 36(7):2295–2300, Apr 2008.

S. Pepenella, K. J. Murphy, and J. J. Hayes. Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure. *Chromosoma*, 123 (1-2):3–13, Mar 2014.

S. Perez-Montero, A. Carbonell, T. Moran, A. Vaquero, and F. Azorin. The embryonic linker histone H1 variant of Drosophila, dBigH1, regulates zygotic genome activation. *Dev Cell*, 26(6):578–590, Sep 2013.

O. Perišić, R. Collepardo-Guevara, and T. Schlick. Modeling studies of chromatin fiber structure as a function of DNA linker length. *J Mol Biol*, 403(5):777–802, Nov 2010.

O. Perišić, S. Portillo-Ledesma, and T. Schlick. Sensitive effect of linker histone binding mode and subtype on chromatin condensation. *Nucleic Acids Res*, 47(10):4948–4957, 06 2019.

S. Poepsel, V. Kasinath, and E. Nogales. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat Struct Mol Biol*, 25(2):154–162, 02 2018.

J. Pointner, J. Persson, P. Prasad, U. Norman-Axelsson, A. Strålfors, O. Khorosjutina, N. Krietenstein, J. P. Svensson, K. Ekwall, and P. Korber. CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. *EMBO J*, 31(23):4388–4403, Nov 2012.

Y. V. Postnikov and M. Bustin. Functional interplay between histone H1 and HMG proteins in chromatin. *Biochim Biophys Acta*, 1859(3):462–467, Mar 2016.

A. Punjani and D. J. Fleet. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol*, 213(2):107702, 06 2021.

A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*, 14(3):290–296, 03 2017.

A. Rada-Iglesias, R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn, and J. Wysocka. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333):279–283, Feb 2011.

V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, 362(6417):219–223, Mar 1993.

S. Rengachari, S. Schilbach, S. Aibara, C. Dienemann, and P. Cramer. Structure of the human Mediator-RNA polymerase II pre-initiation complex. *Nature*, 594(7861): 129–133, 06 2021.

M. A. Ricci, C. Manzo, M. F. Garcia-Parajo, M. Lakadamyali, and M. P. Cosma. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 160(6): 1145–1158, Mar 2015.

V. I. Risca, S. K. Denny, A. F. Straight, and W. J. Greenleaf. Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature*, 541 (7636):237–241, 01 2017.

P. J. Robinson and D. Rhodes. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol*, 16(3):336–343, Jun 2006.

P. J. Robinson, L. Fairall, V. A. Huynh, and D. Rhodes. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A*, 103(17):6506–6511, Apr 2006.

Y. Roulland, K. Ouararhni, M. Naidenov, L. Ramos, M. Shuaib, S. H. Syed, I. N. Lone, R. Boopathi, E. Fontaine, G. Papai, H. Tachiwana, T. Gautier, D. Skoufias, K. Padmanabhan, J. Bednar, H. Kurumizaka, P. Schultz, D. Angelov, A. Hamiche, and S. Dimitrov. The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. *Mol Cell*, 63(4):674–685, 08 2016.

A. Routh, S. Sandin, and D. Rhodes. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A*, 105(26): 8872–8877, Jul 2008.

S. C. Satchwell, H. R. Drew, and A. A. Travers. Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol*, 191(4):659–675, Oct 1986.

T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047):138–141, Jul 2005.

M. P. Scheffer, M. Eltsov, and A. S. Frangakis. Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proc Natl Acad Sci U S A*, 108(41): 16992–16997, Oct 2011.

S. H. Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*, 180(3):519–530, Dec 2012.

S. H. Scheres. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol*, 579:125–157, 2016.

S. Schilbach, M. Hantsche, D. Tegunov, C. Dienemann, C. Wigge, H. Urlaub, and P. Cramer. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature*, 551(7679):204–209, 11 2017.

M. Shimada, W. Y. Chen, T. Nakadai, T. Onikubo, M. Guermah, D. Rhodes, and R. G. Roeder. Gene-Specific H1 Eviction through a Transcriptional Activator→p300→NAP1→H1 Pathway. *Mol Cell*, 74(2):268–283, 04 2019.

R. T. Simpson. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*, 17(25):5524–5531, Dec 1978.

P. J. Skene, A. E. Hernandez, M. Groudine, and S. Henikoff. The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *Elife*, 3:e02042, Apr 2014.

M. Smolle, S. Venkatesh, M. M. Gogol, H. Li, Y. Zhang, L. Florens, M. P. Washburn, and J. L. Workman. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol*, 19(9):884–892, Sep 2012.

F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R. M. Xu, P. Zhu, and G. Li. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science*, 344(6182):376–380, Apr 2014.

E. Stedman and E. Stedman. The basic proteins of cell nuclei. *Phil. Trans. R. Soc. Lond. B*, 235:565–595, 1951.

M. D. Stewart, J. Li, and J. Wong. Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. *Mol Cell Biol*, 25 (7):2525–2538, Apr 2005.

B. D. Strahl and C. D. Allis. The language of covalent histone modifications. *Nature*, 403 (6765):41–45, Jan 2000.

K. Struhl and E. Segal. Determinants of nucleosome positioning. *Nat Struct Mol Biol*, 20 (3):267–273, Mar 2013.

A. Stützer, S. Liokatis, A. Kiesel, D. Schwarzer, R. Sprangers, J. Söding, P. Selenko, and W. Fischle. Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Mol Cell*, 61(2):247–259, Jan 2016.

J. Sun, H. M. Wei, J. Xu, J. F. Chang, Z. Yang, X. Ren, W. W. Lv, L. P. Liu, L. X. Pan, X. Wang, H. H. Qiao, B. Zhu, J. Y. Ji, D. Yan, T. Xie, F. L. Sun, and J. Q. Ni.

Histone H1-mediated epigenetic regulation controls germline stem cell self-renewal by modulating H4K16 acetylation. *Nat Commun*, 6:8856, Nov 2015.

R. K. Suto, M. J. Clarkson, D. J. Tremethick, and K. Luger. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat Struct Biol*, 7(12): 1121–1124, Dec 2000.

H. Tachiwana, W. Kagawa, T. Shiga, A. Osakabe, Y. Miya, K. Saito, Y. Hayashi-Takanaka, T. Oda, M. Sato, S. Y. Park, H. Kimura, and H. Kurumizaka. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature*, 476(7359): 232–235, Jul 2011.

Y. Takizawa, C. H. Ho, H. Tachiwana, H. Matsunami, W. Kobayashi, M. Suzuki, Y. Arimura, T. Hori, T. Fukagawa, M. D. Ohi, M. Wolf, and H. Kurumizaka. Cryo-EM Structures of Centromeric Tri-nucleosomes Containing a Central CENP-A Nucleosome. *Structure*, 28(1):44–53, 01 2020.

P. B. Talbert, K. Ahmad, G. Almouzni, J. Ausio, F. Berger, P. L. Bhalla, W. M. Bonner, W. Z. Cande, B. P. Chadwick, S. W. Chan, G. A. Cross, L. Cui, S. I. Dimitrov, D. Doenecke, J. M. Eirin-Lopez, M. A. Gorovsky, S. B. Hake, B. A. Hamkalo, S. Holec, S. E. Jacobsen, K. Kamieniarz, S. Khochbin, A. G. Ladurner, D. Landsman, J. A. Latham, B. Loppin, H. S. Malik, W. F. Marzluff, J. R. Pehrson, J. Postberg, R. Schneider, M. B. Singh, M. M. Smith, E. Thompson, M. E. Torres-Padilla, D. J. Tremethick, B. M. Turner, J. H. Waterborg, H. Wollmann, R. Yelagandula, B. Zhu, and S. Henikoff. A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin*, 5:7, Jun 2012.

Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher, and D. Lyumkis. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods*, 14(8):793–796, Aug 2017.

J. Taunton, C. A. Hassig, and S. L. Schreiber. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science*, 272(5260):408–411, Apr 1996.

D. Tegunov and P. Cramer. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods*, 16(11):1146–1152, 11 2019.

V. B. Teif, Y. Vainshtein, M. Caudron-Herger, J. P. Mallm, C. Marth, T. Höfer, and K. Rippe. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol*, 19(11):1185–1192, Nov 2012.

A. Thåström, P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, and J. Widom. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J Mol Biol*, 288(2):213–229, Apr 1999.

F. Thoma, T. Koller, and A. Klug. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol*, 83(2 Pt 1):403–427, Nov 1979.

D. J. Tremethick. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell*, 128(4):651–654, Feb 2007.

Y. Tsunaka, N. Kajimura, S. Tate, and K. Morikawa. Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Res*, 33(10):3424–3434, 2005.

M. Turk and W. Baumeister. The promise and the challenges of cryo-electron tomography. *FEBS Lett*, 594(20):3243–3261, 10 2020.

A. Valouev, S. M. Johnson, S. D. Boyd, C. L. Smith, A. Z. Fire, and A. Sidow. Determinants of nucleosome organization in primary human cells. *Nature*, 474(7352):516–520, May 2011.

B. van Steensel and A. S. Belmont. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*, 169(5):780–791, May 2017.

A. Vaquero, M. Scher, D. Lee, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol Cell*, 16(1):93–105, Oct 2004.

S. M. Vos, L. Farnung, A. Linden, H. Urlaub, and P. Cramer. Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat Struct Mol Biol*, 27(7):668–677, 07 2020.

J. P. Wang, Y. Fondufe-Mittendorf, L. Xi, G. F. Tsai, E. Segal, and J. Widom. Preferentially quantized linker DNA lengths in Saccharomyces cerevisiae. *PLoS Comput Biol*, 4(9):e1000175, Sep 2008.

J. D. Watson and F. H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, Apr 1953.

G. Weissenberger, R. J. M. Henderikx, and P. J. Peters. Understanding the invisible hands of sample preparation for cryo-EM. *Nat Methods*, 18(5):463–471, 05 2021.

A. E. White, A. R. Hieb, and K. Luger. A quantitative investigation of linker histone interactions with nucleosomes and chromatin. *Sci Rep*, 6:19122, Jan 2016.

C. L. White, R. K. Suto, and K. Luger. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J*, 20(18):5207–5218, Sep 2001.

J. Widom. Physicochemical studies of the folding of the 100 A nucleosome filament into the 300 A filament. Cation dependence. *J Mol Biol*, 190(3):411–424, Aug 1986.

J. Widom. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc Natl Acad Sci U S A*, 89(3):1095–1099, Feb 1992.

N. Wiechens, V. Singh, T. Gkikopoulos, P. Schofield, S. Rocha, and T. Owen-Hughes. The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLoS Genet*, 12(3):e1005940, Mar 2016.

C. L. Woodcock, L. L. Frado, and J. B. Rattner. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J Cell Biol*, 99(1 Pt 1):42–52, Jul 1984.

C. L. Woodcock, A. I. Skoultchi, and Y. Fan. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res*, 14 (1):17–25, 2006.

J. L. Workman and S Abmayr. Fundamentals of chromatin. 2014.

P. Xu, J. Mahamid, M. Dombrowski, W. Baumeister, A. L. Olins, and D. E. Olins. Interphase epichromatin: last refuge for the 30-nm chromatin fiber? *Chromosoma*, 130 (2-3):91–102, Sep 2021.

K. Yamada, T. D. Frouws, B. Angst, D. J. Fitzgerald, C. DeLuca, K. Schimmele, D. F. Sargent, and T. J. Richmond. Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature*, 472(7344):448–453, Apr 2011.

S. M. Yang, B. J. Kim, L. Norwood Toro, and A. I. Skoultchi. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc Natl Acad Sci U S A*, 110(5):1708–1713, Jan 2013.

X. J. Yang, V. V. Ogryzko, J. Nishikawa, B. H. Howard, and Y. Nakatani. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature*, 382 (6589):319–324, Jul 1996.

Q. Ye and H. J. Worman. Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to Drosophila HP1. *J Biol Chem*, 271(25):14653–14656, Jun 1996.

G. C. Yuan, Y. J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, 309(5734):626–630, Jul 2005.

R. Zhang, J. Erler, and J. Langowski. Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction. *Biophys J*, 112(3):450–459, Feb 2017.

Y. Zhang, Z. Moqtaderi, B. P. Rattner, G. Euskirchen, M. Snyder, J. T. Kadonaga, X. S. Liu, and K. Struhl. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol*, 16(8):847–852, Aug 2009.

E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods*, 18(2):176–185, 02 2021.

B. R. Zhou, H. Feng, H. Kato, L. Dai, Y. Yang, Y. Zhou, and Y. Bai. Structural insights into the histone H1-nucleosome complex. *Proc Natl Acad Sci U S A*, 110(48):19390–19395, Nov 2013.

B. R. Zhou, J. Jiang, H. Feng, R. Ghirlando, T. S. Xiao, and Y. Bai. Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell*, 59(4):628–638, Aug 2015.

B. R. Zhou, H. Feng, R. Ghirlando, S. Li, C. D. Schwieters, and Y. Bai. A Small Number of Residues Can Determine if Linker Histones Are Bound On or Off Dyad in the Chromatosome. *J Mol Biol*, 428(20):3948–3959, 10 2016.

B. R. Zhou, J. Jiang, R. Ghirlando, D. Norouzi, K. N. Sathish Yadav, H. Feng, R. Wang, P. Zhang, V. Zhurkin, and Y. Bai. Revisit of Reconstituted 30-nm Nucleosome Arrays Reveals an Ensemble of Dynamic Structures. *J Mol Biol*, 430(18 Pt B):3093–3110, 09 2018.

B. R. Zhou, K. N. S. Yadav, M. Borgnia, J. Hong, B. Cao, A. L. Olins, D. E. Olins, Y. Bai, and P. Zhang. Atomic resolution cryo-EM structure of a native-like CENP-A nucleosome aided by an antibody fragment. *Nat Commun*, 10(1):2301, 05 2019.

B. R. Zhou, H. Feng, S. Kale, T. Fox, H. Khant, N. de Val, R. Ghirlando, A. R. Panchenko, and Y. Bai. Distinct Structures and Dynamics of Chromatosomes with Different Human Linker Histone Isoforms. *Mol Cell*, 81(1):166–182, 01 2021a.

K. Zhou, M. Gebala, D. Woods, K. Sundararajan, G. Edwards, D. Krzizike, J. Wereszczynski, A. F. Straight, and K. Luger. CENP-N promotes the compaction of centromeric chromatin. *bioRxiv*, 2021b. doi: 10.1101/2021.06.14.448351. URL `https://www.biorxiv.org/content/early/2021/06/30/2021.06.14.448351`.

V. B. Zhurkin and D. Norouzi. Topological polymorphism of nucleosome fibers and folding of chromatin. *Biophys J*, 120(4):577–585, 02 2021.

Supplementary Materials

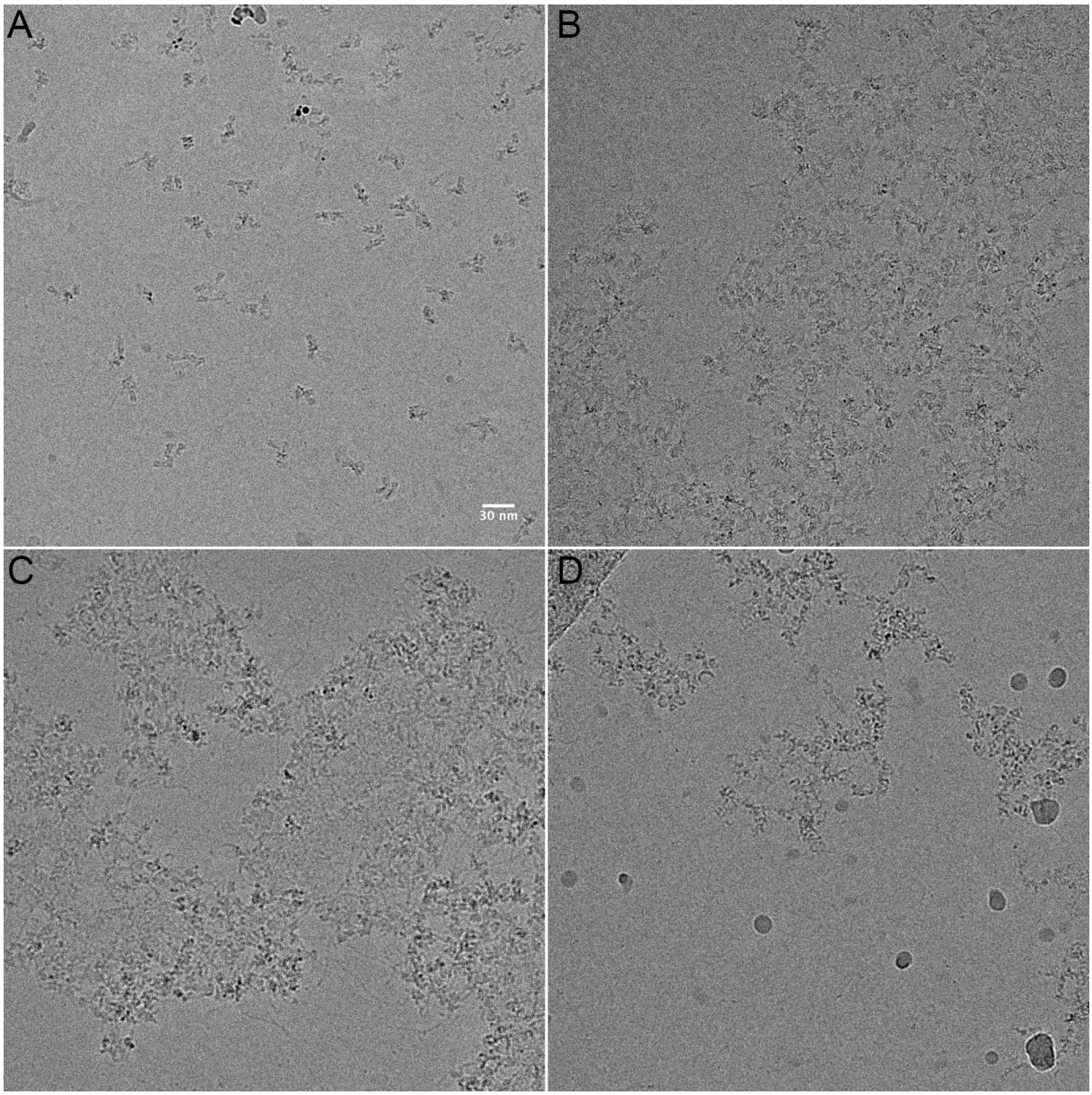

Figure A.1: Cryo-screening of 4x188 arrays with different NaCl concentrations. In the beginning of the project different NRL were tested, including 188 bp for which a template with 12 nucleosomes already was available in the lab. Nucleosome arrays were reconstituted and different sodium chloride concentrations (**(A)** 0 mM, **(B)** 50 mM, **(C)** 100 mM, **(D)** 150 mM) were tested for cryo-EM sample prepartion and screened in a Glacios 200 kV transmission electron microscope. Initial screening results showed well separated particles for sample without any salt **(A)**. These conditions were also used for the previously published cryo-EM structure of the 12x177 and 12x188 (Song et al., 2014). However, during later preparations of 4x177, 4x187, 4x197 and 4x207 a large fraction of cryo-EM grids prepared under these conditions also presented aggregated or clustered particles as in **(B)**, **(C)** and **(D)**, indicating that suitable sample might also be possible with higher salt concentrations.

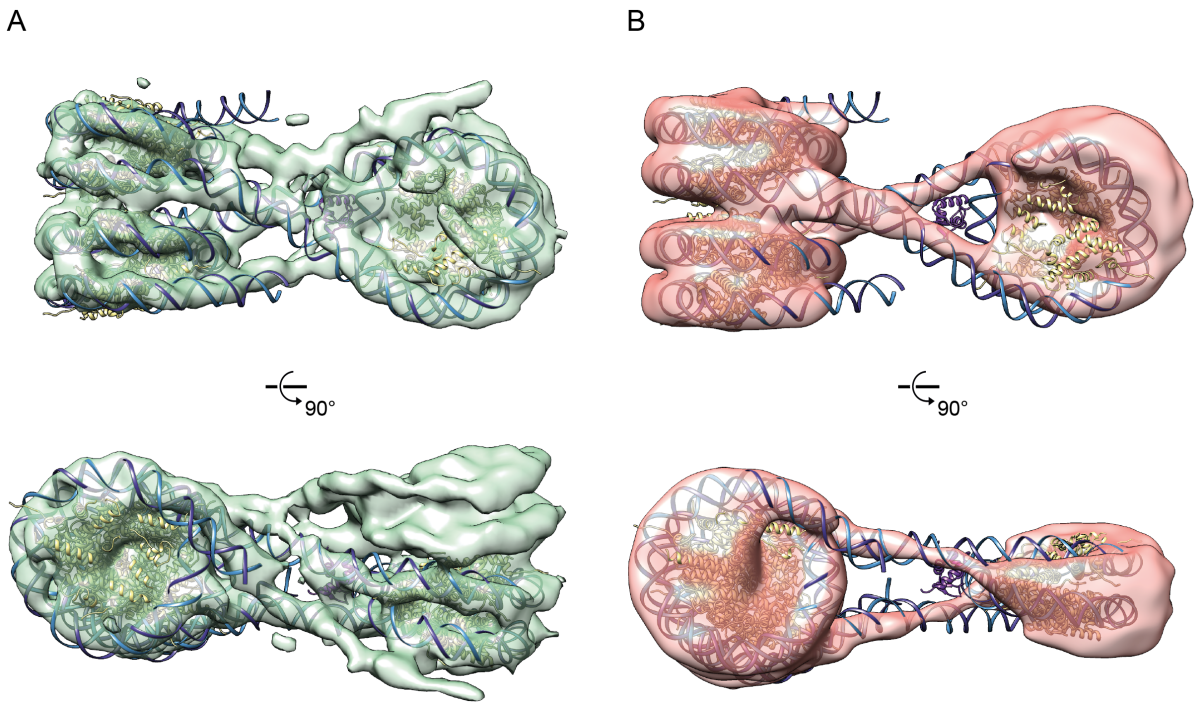

Figure A.2: The structures of compacted nucleosome arrays. The cryo-EM structures of the **(A)** tetranucleosome unit of the 4x177 compacted with H1 (green) and **(B)** 3x177 trinuc compacted with Mg2+ (red) fitted with the 4x177 trinucleosome core determined in this work (DNA in blue, core histones in khaki, H1 in purple).